8-2017

The Tennessee Educator Acceleration Model and the Teacher Instructional Growth for Effectiveness and Results Model on Measures of Teacher Effectiveness: A Comparative Study

Nichole R. Morris
East Tennessee State University

The Tennessee Educator Acceleration Model and the Teacher Instructional Growth for Effectiveness and Results Model on Measures of Teacher Effectiveness: A Comparative Study

_____

A dissertation

presented to

the faculty of the Department of Educational Leadership and Policy Analysis

East Tennessee State University

In partial fulfillment

of the requirements for the degree

Doctor of Education in Educational Leadership

_____

by

Nichole Renee Morris

August 2017

_____

Dr. Pamela Scott, Chair

Dr. Donald Good

Dr. Laura Robertson

Dr. Stephanie Tweed

Keywords: Teacher Evaluations, TEAM, TIGER, TVAAS, Teacher Effectiveness, Value-added, Observations

ABSTRACT

The Tennessee Educator Acceleration Model and the Teacher Instructional Growth for Effectiveness and Results Model on Measures of Teacher Effectiveness: A Comparative Study

by

Nichole Morris

The purpose of this quantitative study was to compare measures of teacher effectiveness between two different Tennessee teacher evaluation models. The two teacher evaluation models compared were the Tennessee Educator Acceleration Model (TEAM) and the Teacher Instructional Growth for Effectiveness and Results Model (TIGER). The measures of teacher effectiveness used were final observation scores and individual value-added growth (TVAAS) scores. The relationship between observation scores and growth scores were also compared for the two different evaluation models. There were four guiding research questions and corresponding null hypotheses. Data were provided from participating school systems. The sample for this study included 230 TIGER teachers from 3 TIGER districts and 2,389 TEAM teachers from 9 TEAM districts. Independent samples t-tests and Pearson correlations were computed to investigate the research questions. All data were analyzed using a 0.05 significance level. Findings from the data indicate that teachers who were evaluated using the TIGER model had statistically higher TVAAS scores and final observation scores when compared with TEAM teachers. Results also indicated a significant positive relationship between observation scores and TVAAS scores for both TEAM and TIGER teachers. Findings indicate a stronger relationship for the TIGER teachers than for the TEAM teachers.

DEDICATION

This dissertation is dedicated to my family. To Jacob, my husband, I could not have done this without your support and encouragement. Thank you for everything. You are amazing. I love you. To my son, Samuel, I love you more than I can ever explain. I want you to always remember that learning never ends. With hard work and perseverance, anything is possible.

ACKNOWLEDGEMENTS

I want to acknowledge and thank Dr. Scott for being a supportive and helpful dissertation chair. Thank you for guiding me through this process. I also want to acknowledge Dr. Good. Thank you for being available to answer all of my statistics questions. Thank you also for your suggestions about ways to improve my work. I also want to recognize Dr. Robertson and Dr. Tweed for serving on my dissertation committee. Dr. Tweed, I appreciate all of your APA help and positive critique throughout this process. Dr. Robertson, thank you for advice and constant encouragement.

I also want to acknowledge Heather Boegemann. Without you, this would not have been possible. From helping me with my practicum to answering any TIGER or evaluation question that I had, I am forever grateful.

I also want to acknowledge my fellow students and friends in this ELPA program. To Amy Horton, Kari Eubanks, Aleeta Shaw, and Kelli Barnett, thank you for all of your support throughout this process. You are all amazing.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1
INTRODUCTION

In 2009, the American Recovery and Reinvestment Act was passed in the United States. As a part of this Act, the Race to the Top (RTTT) program was created in order to reform the education system and to help make students more prepared for college and work. There are four main goals of the Race to the Top program. The first is for states to adopt more rigorous standards and assessments in order for students to be ready for college or the workplace. The second is for states to create data systems that help measure student growth and to enable educators to discern areas for improvement. The third is to find ways to support and develop teachers as well as develop incentives to recruit and retain high performing teachers. The last goal is to improve the lowest performing schools (United States Department of Education, 2016). States were challenged to design reform plans to meet these four goals. States that were successful in their design were eligible for funds to help support their efforts.

Tennessee was one of the first states to apply for and receive Race to the Top Funds. As a part of this process, Tennessee passed the First to the Top Act of 2010, creating the Teacher Evaluation Advisory Committee, which made changes to the teacher evaluation system (Tennessee First to the Top Act, 2010). The goal of these changes to the teacher evaluation system was to increase student achievement (Tennessee Department of Education, 2012). The purposes of teacher evaluations, as identified by the Tennessee State Board of Education (2015), are to find and develop instructional practices that contribute to student achievement, inform human capital decisions such as hiring, firing, and promotions, and to differentiate teachers in to

specified effectiveness groups. These effectiveness groups range from significantly above expectations to significantly below expectations.

Beginning with the 2011-2012 school year, changes were implemented regarding how Tennessee teachers were evaluated. The First to the Top Act based teacher evaluations on three main components: student achievement growth measures, value-added measures, and observations. Fifty percent of a teacher's evaluation would come from student achievement data. Of this 50%, 35% would come from a measure of student growth from the Tennessee Value-Added Assessment System (TVAAS), and the other 15% would come from other measures of student growth as decided on by the teacher and administrator. The remaining 50% of a teacher's evaluation would come from other qualitative measures such as observations, conferences, and a review of previous work and evaluations (Tennessee First to the Top Act, 2010). For teachers who teach a subject that is not tested, 25%of their evaluation would come from school wide TVAAS, 15% from other measures of student growth as decided by the teacher and the principal, and 60% of their evaluation would come from qualitative measures. These percentages are different for the 2015- 2017 school years due to the implementation of new standardized tests, but are planned to return to the original percentages after full implementation during the 2017-2018 school year (Tennessee Department of Education, 2016b). Tennessee's teacher evaluations are consistent with the current trend in teacher evaluation models which combine observation data with value-added data (Hill & Grossman, 2013). At least 41 states use an evaluation model that incorporates some combination of value-added measure, observations, and other artifacts such as surveys and lesson plans (Hull, 2013).

With guidance from the Teacher Evaluation Advisory Committee, the Tennessee Department of Education adopted the Tennessee Educator Acceleration Model (TEAM) as the

official teacher evaluation model for the state of Tennessee. However, there are four alternate teacher evaluation models that have been approved for use by the Tennessee State Board of Education (Tennessee State Board of Education, 2015). They are the Teacher Instructional Growth for Effectiveness and Results Model (TIGER), Project COACH, Teacher Effectiveness Measure (TEM), and The Achievement Framework for Excellent Teaching (AFET). In order to be approved as a Tennessee teacher evaluation model, the model must include a way to measure observations in the four domains of planning, environment, instruction, and professionalism. Approved models must also involve conferencing and reflections on past performance. School systems have the option of determining which model to use when evaluating teachers.

TEAM is the official teacher evaluation model for the state of Tennessee. The TIGER model is the most used of the four alternative teacher evaluation models. Both models use the same TEAM rubric to assess teacher performance through classroom observations. How they differ is in the structure of the observations, the use of artifacts, and final observation scoring. With the TEAM model, each observation focuses on a specific domain from the TEAM rubric. The TIGER model focuses on all domains from the TEAM rubric for each observation. With the TEAM model, a teacher's final observation score is an average of all of the scores from the observations throughout the year. With the TIGER model, a teacher's final observation score is determined from observation scores throughout the year in combination with evidence or artifacts. In the TIGER model, teachers can provide artifacts such as student work, assessments, and lesson plans to serve as evidence for success on an indicator from the TEAM rubric. The administrator will consider the observation data and evidence provided to assign a score for each indicator on the rubric (Teacher Instructional Growth for Effectiveness and Results, 2016; Tennessee Department of Education, 2016b).

Because systems have the freedom to select an approved evaluation model, it is important to understand whether the evaluation model has any impact on measures of teacher effectiveness and student learning. This study examined the influence of the teacher evaluation model on teacher success. The purpose of this study was to compare two different Tennessee teacher evaluation models that are being used on measures of teacher effectiveness. The two models were the Tennessee Educator Acceleration Model (TEAM) and the Teacher Instructional Growth for Effectiveness and Results Model (TIGER), which are the two most commonly used evaluation models in Tennessee (Tennessee Department of Education, 2015). While both models have similar components, there are differences that have the possibility of impacting teacher performance.

Statement of the Problem

There is a lack of research regarding the TEAM and TIGER evaluation models. Much of the available research focuses on teacher perceptions of the evaluation model and process, typically using surveys to collect data (e.g. Bogart, 2013; Bryant, 2013; Ehlert, Pepper, Parsons, Burns, & Springer, 2013; New Teacher Center, 2013; Pepper, Ehlert, Parsons, Stahlheber, & Burns, 2015). Also, a majority of existing research, with exceptions including Bogart (2013) and Bryant (2013), has been performed by the Tennessee Department of Education or an affiliate. Of the two evaluation models, a small amount of academic research has been conducted with the TEAM model. Both Mason (2015) and Davis (2014) investigated the relationship between observation scores and student achievement for teachers who were evaluated using the TEAM model. Results were conflicting from both of these studies. There is an overall lack of published research regarding the two evaluation models. No available study could be located that specifically compared the two models in terms of teacher effectiveness. It is important to

understand if the evaluation model has any impact on teacher effectiveness as 50% of a teacher's evaluation score comes from observations and 35% comes from value-added scores for teachers who teach a testable subject (Tennessee Department of Education, 2016b). It is also important because human resource decisions may be based upon teacher effectiveness ratings.

Purpose Statement

The purpose of this quantitative, causal comparative study was to examine whether there was a difference in measures of teacher effectiveness between two different Tennessee teacher evaluation models: the Tennessee Educator Acceleration Model (TEAM) and the Teacher Instructional Growth for Effectiveness and Results (TIGER) model. The dependent variable of teacher effectiveness was measured by final observation scores from the TEAM rubric and value-added scores (TVAAS). The independent variable was the teacher evaluation model. The relationship between observation scores and value-added scores was also compared between the two different evaluation models.

Research Questions

The purpose of this study was to compare the TEAM and TIGER Tennessee teacher evaluation models on measures of teacher effectiveness. The following research questions guided this study.

Research Question 1

Is there a significant difference in overall teacher effectiveness as measured by TVAAS between teachers who are evaluated with the TEAM model and teachers evaluated with the TIGER model?

Research Question 2

Is there a significant difference in final principal observation scores between teachers who are evaluated with the TEAM model and teachers evaluated with the TIGER model?

Research Question 3

Is there a significant relationship between teachers' value-added growth scores (TVAAS) and teachers' final observation scores for teachers who are evaluated with the TEAM model?

Research Question 4

Is there a significant relationship between teachers' value-added growth scores (TVAAS) and teachers' final observation scores for teachers who are evaluated with the TIGER model?

Significance of the Study

The purpose of this study was to compare measures of teacher effectiveness between two different Tennessee teacher evaluation models. There is currently a limited amount of published research on the TEAM and TIGER evaluation models. There is also a lack of research on the relationship between teacher evaluation models and measures of teacher effectiveness. Most of the available research focuses on specific aspects of the evaluation model such as observations (e.g. Davis, 2104; Garret & Steinberg, 2015; Grossman, Cohen, Ronfeldt, & Brown, 2014; Jacob & Lefgren, 2008) or value-added measures (e.g. Ballou, Sanders, & Wright, 2004; Staiger & Kane, 2014). There are some studies that look at the relationship between value-added measures and observations (e.g. Harris, Ingle, & Rutledge, 2014; Kane, Taylor, Tyler, & Wooten, 2011), but little available research that investigates the relationship between evaluation model and

teacher effectiveness. This research study could add to the body of research on teacher evaluation models and teacher effectiveness.

This study is significant because it could influence understanding about the components of effective teacher evaluation models. While the TEAM and TIGER models share the same rubric, there are differences in the focus of observations, scoring, and the use of evidence and artifacts. These differences could potentially impact teacher success and student achievement. Results of this study could have implications for future evaluation research.

This study is also significant because evaluations relate to human capital decisions. Because evaluations are used by many school systems to inform such decisions as hiring, firing, and promotion of teachers, it is important to gain a better understanding of the relationship between measures of teacher effectiveness and the teacher evaluation model.

Definitions of Terms

The following definitions explain the meaning of words that pertain to the content included in this study.

End of Course (EOC) Exams: End of Course exams are the Tennessee assessments given to high school students in the subjects of English I, English II, English III, Algebra, Geometry, Algebra II (or Integrated Math), United States History, Chemistry, and Biology (Tennessee Department of Education, 2016a).

Race to the Top (RTTT): Race to the top is a government initiative to generate educational reform in four areas: adopting curriculum standards and assessments for students to be successful, the creation of a data system to measure student achievement, the development of ways to recruit,

support, and retain effective teachers, and to raise the achievement of the lowest performing schools (United States Department of Education, 2016).

Standards Based Observation: A standards based observation is an observation that uses a rubric that is based on a standard set of effectiveness characteristics (Henemen, Milanowski, Kimball, & Odden, 2006).

Teacher Effectiveness: Teacher effectiveness is the ability to improve student learning (Danielson, 2007).

Teacher Instructional Growth for Effectiveness and Results Model (TIGER): The Teacher Instructional Growth for Effectiveness and Results model is one of the alternative teacher evaluation models for the state of Tennessee (Teacher Instructional Growth for Effectiveness and Results, 2016).

Tennessee Comprehensive Assessment Program (TCAP): The Tennessee Comprehensive Assessment Program is the state of Tennessee's standardized assessment program that is used to assess student learning in the subjects of math, science, social studies, and language arts (Tennessee Department of Education, 2016a).

Tennessee Educator Acceleration Model (TEAM): The Tennessee Educator Acceleration Model is the official teacher evaluation model for the state of Tennessee (Tennessee Department of Education, 2016b).

Tennessee Value-Added Assessment System (TVAAS): The Tennessee Value Added Assessment System is the value added component of the Tennessee teacher evaluation model. TVAAS measures student growth and the impact that schools and teachers have on student achievement

(Tennessee Department of Education, 2016b). Both the TIGER and TEAM models incorporate a teacher's TVAAS score in the computation to determine a teacher's effectiveness rating.

Value-added: Value-added is a system used to measure student learning growth. Value-added systems also provide an estimation of teacher, school, and school system effects on a student's growth (Tennessee Code Annotated, 2016).

Limitations

The use of TVAAS was a limitation to the research study. There is a perception among some researchers that value-added methods are not reliable or valid ways to measure teacher effectiveness because there are components of teacher effectiveness that cannot be measured in an algorithm (Amrein-Beardsley, 2008; Papay, 2012). There is also concern about the reliability of the TCAP and EOC assessments to accurately measure student learning (Baker & Xu, 1995; SCORE, 2014). Scores from these assessments are used to determine a teacher's TVAAS. Because the state of Tennessee has chosen to use TVAAS to measure teacher effectiveness, it was used to represent teacher effectiveness in this study.

The use of final observation scores was a potential limitation. There is some published research that suggests that observation data is not always reliable because ratings on observation rubrics can be subjective (Papay, 2012). Because the state of Tennessee uses final observation scores as part of the measurement of teacher effectiveness, the scores were used as a part of this study. Tennessee requires all evaluators to be trained in using and evaluating teachers with the TEAM rubric. Evaluators must pass an online certification test in order to perform teacher observations (Tennessee Department of Education, 2016b).

Another limitation was the sampling. Convenience sampling was used to identify participants for this study. To complete this study, pre-existing data was needed, which could only be provided by individual school systems. The researcher attempted to contact every TEAM and TIGER school to request data, but had to use school systems that were willing to participate. Random sampling was not possible due to the nature of the data. Because the researcher had to rely on willing participants, there is no way to ensure that the schools systems that participated were representative of the population being studied (Creswell, 2002).

Delimitations

This study was delimited to the 2014-2015 school year. Many issues surrounded state standardized testing in the 2015-2016 school year such as technology issues, late mailing of the print version of the tests, cancelling the contract with the testing vendor, and consequently not requiring schools to complete testing, so it was decided to use data from the 2014-2015 school year. It was the concern of the researcher that the data from the 2015-2016 school year would present too many reliability and validity concerns.

The study was also delimited to the TEAM and TIGER Tennessee teacher evaluation models. There are three other approved alternative teacher evaluation models that were not considered in this study. TEAM was chosen because it is the official evaluation model for the state of Tennessee, and TIGER was chosen because it is most used model of the approved alternative models.

This study was delimited to Tennessee teachers who taught a tested subject. However, many teachers do not teach a tested subject. Therefore, they were not included in this study because they would not have an individual TVAAS score to incorporate into the data.

Overview of the Study

The purpose of this study was to compare measures of teacher effectiveness between the TEAM and TIGER teacher evaluation models. This study will be presented in five chapters. Chapter 1 provides an introduction, purpose of the study, research questions, significance of the study, limitations and delimitations, and definitions of important terms. Chapter 2 is a review of relevant literature that focuses on teacher effectiveness, teacher evaluations, teacher observations, and value-added measures. Chapter 3 describes the methodology and data collection method used for this study. Chapter 4 contains the data and data analysis for this study. Chapter 5 contains a summary and recommendations for future areas of research.

CHAPTER 2

REVIEW OF LITERATURE

Since the 1990s, accountability has been a large focus of educational reform. School accountability is the, "process of evaluating school performance of the basis of student performance measures" (Figlio & Loeb, 2011, p. 384). Standardized testing of students to determine academic growth has been a major component of the accountability movement. Government initiatives such as the No Child Left Behind Act of 2001 (2002) and the more recent Every Student Succeeds Act (2015) aimed to ensure learning for all students while holding teachers and schools responsible for achievement and success.

Some of the concern regarding accountability and student performance has arisen from lackluster performance as a nation on international standardized tests. The Program for International Student Assessment (PISA) is given every three years to 15-year-old students in order to measure knowledge in the subjects of math, science, and reading. In 2012, the United States ranked 36th for average math literacy score, 27th for average science literacy score, and 24th for average reading literacy score (Kelly et al., 2013). United States results on the Trends in International Mathematics and Science Study (TIMSS) were better. The TIMSS study is performed every four years to assess student achievement in math and science. In 2011, the United States ranked 11th in fourth grade math and 9th in eighth grade math (Mullis, Martin, Foy, & Arora, 2012). In science, fourth grade students ranked 7th and eighth grade students ranked 10th internationally (Martin, Mullis, Foy, & Stanco, 2012). There is a growing concern due to these and other similar results that students are not being effectively prepared to be competitive with their international counterparts, and the accountability movement has stressed the

importance of increasing achievement for all students while recognizing the significant impact of teachers on student learning.

With the current state of teacher evaluations, teachers are held accountable for the academic growth and progress made by their students. Teachers have been shown to have a large impact on student learning (Aaronson, Barrow, & Sander, 2007; Hanushek, 1992). It is therefore important to study different aspects that can impact teacher effectiveness and student achievement. The purpose of this study was to investigate the relationship between a teacher evaluation model and measures of teacher effectiveness. Specifically, this study was a comparison of the TEAM and TIGER Tennessee teacher evaluation models in terms of teacher effectiveness as measured by value-added growth scores and observation scores. This study also compared the relationship between value-added scores and observation scores for the two evaluation models. This literature review will look at prior research regarding teacher effectiveness, teacher evaluations, standards based observations, and value-added growth measures.

Framework for Effective Teaching

In 1996, Danielson published a framework to identify research supported educational practices that increase student learning. Danielson's framework acknowledged that teaching is a complex task that involves many different activities. As a result, the framework centered effective teaching around four domains: planning and preparation, classroom environment, instruction, and professionalism. Within each domain, there are components which address essential characteristics of effective teaching, and each component has elements which describe the important aspects of each component. These elements, "define what teachers should know

and be able to do in the exercise of their profession" (Danielson, 2007, p. 1). Many of the teaching practices described by Danielson are incorporated into current teacher observation rubrics, including the TEAM rubric, which are used to evaluate teachers.

The first domain of Danielson's 2007 framework is planning and preparation. This domain focuses on how a teacher prepares, designs, and organizes lessons for student learning. Within this domain, an effective teacher is knowledgeable about his or her students, the subject area, learning strategies, content standards, and educational resources. An effective teacher should use both formative and summative assessments to determine a student's strengths and weaknesses, and then match instructional strategies in order to meet student needs. Lessons are designed in a way that allow for students to move from simple to complex ideas. Critical thinking and problem solving skills are supported and developed in students.

The second domain in Danielson's 2007 framework focuses on the classroom environment. "The components of Domain 2 establish a comfortable and respectful classroom environment that cultivates a culture for learning and creates a safe place for risk taking" (p. 28). An effective teacher in this domain develops positive relationships with his or her students. Students feel cared for, and as a result trust that they can take risks. All students are held to high expectations. Little instructional time is wasted, and the teacher has strong classroom management and organization skills. A successful teacher organizes the classroom in a way that is conducive to learning and activities (Danielson, 2007).

Danielson's 2007 third domain for effective teaching is instruction. An effective teacher in this domain successfully carries out the plans and lessons created in domain one. "And the heart of Domain 3 is engaging students in learning; all the other aspects of the framework serve

the purpose of engagement, because it is engagement that ensures learning" (p. 77). A successful teacher with this component is one who can explain directions, objectives, and information clearly for students and in a way that they understand. Higher order thinking questions are designed to facilitate complex thought and discussion. Students are engaged and involved in the learning process. Formative and summative assessments are used to provide feedback in order to help students be successful. An effective teacher can read student cues to make changes to the lesson as needed and offer support to students when necessary (Danielson, 2007).

Danielson's 2007 final domain in the framework is professionalism. This domain deals with a teacher's interactions inside and outside of the classroom with parents and the community. An effective teacher keeps parents informed about what students are learning, what is happening in the classroom, student progress, student successes as well as concerns about students. This component is important because successful communications with parents can lead to supportive and positive relationships. Professionalism also refers to a teacher's involvement in professional growth activities such as self-reflection, collaboration with colleagues, and professional development opportunities. An effective teacher in this domain is a model of integrity, respect, and responsibility in terms of their interactions with colleagues and students.

Danielson's 2007 framework defines and describes the characteristics of effective teaching. The components and characteristics in this framework are included in the TEAM observation rubric currently in use in the state of Tennessee to evaluate teachers for both the TEAM and TIGER models.

Teacher Effectiveness

The definition of teacher effectiveness has varied and changed over time. The current focus on accountability centers the definition around student achievement. The United States Department of Education (2009) defines a highly effective teacher as, "a teacher whose students achieve high rates (*e.g*, one and one-half grade levels in an academic year) of student growth" (p. 12). While this definition focuses mainly on value-added and achievement growth, others have defined effectiveness as a teacher's ability to increase student learning based on student performance (Darling-Hammond, 2000; Munoz & Chang, 2007; Stronge, Ward, Tucker, & Hindman, 2008).

The characteristics of effective teaching have been a growing area of research interest. Teachers have a direct impact on student achievement, and much research has been devoted to what qualities, characteristics, and practices increase teacher effectiveness. Research has been focused on two main areas: the qualifications of effective teachers and the instructional strategies and methods used by effective teachers. In most of the available research, teacher effectiveness has been determined based upon student achievement and performance on learning measures (e.g. Chingos & Peterson, 2011; Decristan et al., 2015; Grant, Stronge, & Xu, 2013; Munoz, Prather, & Stronge, 2011; Stronge, Ward, & Grant, 2011).

Much of the available research on the instructional strategies of effective teachers has focused on what differentiates an effective teacher from a less effective teacher. Stronge et al. (2011) investigated these differences with fifth grade teachers from three different school systems within one state. The researchers compared effectiveness based upon four different components: instructional delivery, the use of assessment, learning environment, and the

characteristics of the teacher. Teacher effectiveness was determined from value-added scores on state achievement tests. The researchers reported that strong classroom management, well organized classrooms, fostering positive relationships with students, and allowing students to have increased responsibility were all factors that had a positive significant impact on student achievement and teacher effectiveness.

A 2008 study by Stronge et al. compared high performing and low performing third grade teachers in Virginia to see if there were any differences in classroom practices. Value-added scores were used to place teachers in either the high or low performing category. Observations were used to determine any differences. The researchers discovered that effective teachers tended to differentiate lessons more as well as incorporate more complex thought and tasks into their lessons. Effective teachers had high expectations for their students. The classrooms of effective teachers were organized, efficient, and generally had less off-task student behavior. Successful teachers had a classroom environment that was built on fairness and respect. A similar study was performed in 2011 by Kane et al. The researchers used observation data from teachers in Cincinnati specifically looking for practices that increased student achievement. The researchers found that in math classrooms, effective teachers had strong classroom management. They also reported that in reading classrooms, successful teachers effectively used questions and discussions to facilitate learning.

Tyler, Taylor, Kane, and Wooten (2010) reviewed teacher evaluation data from Cincinnati teachers to determine what effective teachers did to foster student learning. Cincinnati teachers are evaluated using the Teacher Evaluation System (TES), which is based on the work of Danielson. The researchers specifically focused on the domains of learning environment and teaching from the rubric. The learning environment domain focuses on creating a caring and

supportive classroom, classroom routines, and classroom management. The teaching domain focuses on instruction, high expectations, objectives, and content knowledge. Tyler et al. (2010) found that teachers who scored higher in the area of classroom environment when compared to instructional practices had higher student achievement in both math and reading. This research, along with the research of others (Kane et al., 2011; Stronge et al., 2008; Stronge et al., 2011) highlights the idea that effective classroom teachers have high levels of classroom management.

Hattie (2012) extensively reviewed the literature on effective and successful teaching and identified five different characteristics of expert or successful teachers. The first is that effective teachers are able to recognize the best ways to represent and model the content in the subject that they teach. They are able to organize the content in a way that facilitates learning and deep understanding as well as remediate misconceptions. The next characteristic of effective teaching as identified by Hattie is the ability to create a classroom environment that supports the development of trusting relationships. Students feel comfortable and safe in taking risks and sharing ideas. Providing feedback to students is the third characteristic of effective teachers. Effective teachers can recognize when a student needs extra support as well as provide help in a quick and timely manner. Hattie also found that effective teachers have high expectations for all students. They believe that all students can succeed and learn. The final characteristic of effective teachers is the ability to influence student outcomes through challenge and engagement. The characteristics identified by Hattie are evident in Danielson's 2007 work.

Grossman et al. (2010) investigated what differentiated successful and less successful teachers specifically in middle school language arts classrooms. Teachers were identified as successful based upon value-added scores. The researchers observed that high performing teachers provided specific strategies and methods to students to help them approach learning

activities. They also had lessons that were authentic and extended and applied to other content areas. Effective teachers tended to provide feedback to students more frequently as well as scaffold difficult content for students. Modeling and the use of multiple representations of new content was evident in the lessons of effective teachers. Specifically, in terms of language arts, successful teachers focused more on writing, speaking, and research skills over reading and literature skills.

Decristan et al. (2015) studied teaching practices that were effective in increasing science achievement. The researchers were interested in the effects of formative assessment, instructional supports, and classroom management on student science learning. The researchers described instructional supports as activating prior knowledge, making connections, critical thinking and problem solving. The study was performed with 28 teachers in Germany. Student learning was measured with a test that consisted of multiple choice and open response questions. Results of the study indicated that formative assessment, instructional support, and classroom management all had a positive impact on student learning of science content.

Mangiante (2011) examined practices of effective teachers that increased achievement for low income, minority, students. Many of the practices identified by Mangiante to help minority students were also identified by other researchers as effective practices for all students. These include high expectations, the use of scaffolding, clear goals, feedback, the incorporation of formative assessments, and routines. Mangiante identified additional strategies used by effective teachers to help minority students such as making connections to the community and culture of the students, engaging in self-reflection and assessment, developing trusting relationships, and holding the belief that these students can succeed.

Grant et al. (2013) studied award winning teachers in the United States and China. They were interested in what the similarities and differences were between these recognized effective teachers. In both the United States and China, effective teachers used a variety of instructional activities. These teachers did not rely on only one instructional method. Effective teachers incorporated activities and lessons that focused on the areas of knowledge, comprehension, application, and analysis from Bloom's Taxonomy. Like Stronge et al. (2011), the researchers discovered that effective teachers had strong classroom management. Students were often engaged and on task. There were multiple differences between effective Chinese and American teachers. Award winning American teachers also incorporated differentiation, creativity, and the exploration of ideas in their lessons. They had positive interactions with their students and would often allow students to direct or control their own learning.

The research regarding the instructional practices of effective teachers has provided an extensive list of strategies that can be developed and incorporated by teachers to increase the learning and achievement of students. The other area of research on teacher effectiveness has focused instead on the qualifications of effective teachers. Specifically, these studies have attempted to determine if characteristics such as years of experience or educational level have an impact on student achievement and teacher success (e.g. Chingos & Peterson, 2011; Darling-Hammond, 2000; Munoz et al., 2011; Shuls & Trivitt, 2015).

Two similar studies investigated teacher characteristics that positively impacted achievement as measured by value-added scores. Munoz et al. (2011) investigated teachers from one school system in Kentucky. They specifically investigated whether the number of students in a class, a teacher's years of experience, a teacher's highest level of education, a teacher's rank, race, or gender had any impact on reading achievement. They found that there was a significant

difference in reading achievement with years of experience. Chingos and Peterson (2011) performed a similar study in Florida, but found somewhat conflicting results. They found that college major, advanced degrees, and selectivity of college or university had no significant impact on teacher effectiveness. They did discover a positive relationship between years of experience and achievement, but this did not hold up over time. Positive relationships were also identified between National Board Certification and achievement as well as scores on teacher certification exams and achievement. While both of these studies had different results, both identified a positive relationship between years of experience and teacher effectiveness.

Darling-Hammond (2000) used surveys and student achievement data to research teacher qualifications and their relationship to effectiveness and achievement. The researcher found that there was a positive correlation between highly qualified teachers and student achievement as well as a positive correlation with achievement when teachers had a degree in their career field. A teacher's level of education had a positive relationship to student achievement, but was found to not be as strong as the other correlations. Shuls and Trivitt (2015), in a similar study, investigated teacher certification and teacher scores on teacher licensure exams to see if there was any impact on effectiveness. The researchers studied Arkansas teachers in grades three through eight. They considered whether a teacher was traditionally or alternatively certified, and they used scores from Praxis exams to investigate. Results indicated that teachers who scored higher on the teacher licensure exams tended to perform better at raising student achievement. Shuls and Trivitt also found that alternatively certified teachers tended to perform lower in math and language arts, but this did not hold up when the data was restricted only to teachers who taught a tested subject.

A 2007 study by Munoz and Chang of high school reading teachers in Kentucky investigated the impact of a teacher's education, years of experience, or race on teacher effectiveness. The researchers found that these characteristics had no significant impact on teacher effectiveness or student performance. The results of the research on teacher qualifications and teacher effectiveness have produced mixed and varied results. There is little consistency in the research to support the idea that teacher qualifications have much impact on student learning.

Teacher Evaluations

History of Teacher Evaluations

The function and implementation of teacher evaluations has varied greatly throughout history. There have been many changes to teacher evaluation systems based on the events, values, and concerns at the time. While the methods may be different, there has been some form of evaluation or assessment of teaching since the 1600s. Concepts of evaluating teacher effectiveness are evident as early as 1659 when Hoole published a pamphlet about educational strategies and teaching. He suggested that an effective teacher is one who is viewed positively in the community and is respected by parents. He also recommended that teachers should be models of moral behavior. According to Hoole, a successful teacher should maintain order in the classroom as well as correct students when wrong but not to the extent of embarrassment.

Early teacher evaluations had very little regulations or criteria in order to judge effective teaching. In the 1700s, most of the supervision of teachers was performed by members of the clergy or other religious leaders. Most of the human resource decisions were based simply on opinion and perceptions (Marzano, Frontier, & Livingston, 2011). Even into the early 1900s,

successful teachers were graded on attributes such as appearance, integrity, and speech (Shinkfield & Stufflebeam, 1995). Early teacher evaluations were rarely based on research or qualities of effective teaching. It was not until the early 20th century that teaching and evaluations received more focus and importance.

The efficiency movement or scientific management movement began around 1915 and extended into the 1920s based on the work of Cubberly and Bobbit (Callahan, 1962). The focus was on improving student learning and teacher performance based on methods that produced the best results. Different teaching methods and strategies were to be used and assessed. Those strategies that did not produce the desired results were to be eliminated (Cubberly, 1916). This would help to improve efficiency and effectiveness. Teachers were often measured or evaluated using efficiency scales. One example of an efficiency scale used by a school system rated teachers in the areas of influence on student, teaching ability, discipline, scholarship, promptness, teacher growth, and results (Callahan, 1962). Another efficiency scale rated teachers in the areas of physical environment, teacher personality, adaptability, adherence to school policies, cooperation, relations with students, classroom management, teaching skills, professionalism, and overall impression (Cubberly, 1916). Observations were mainly performed by principals or administrators. Cubberly recommended using multiple observers for each teacher to get a more accurate picture of a teacher's efficiency.

With the scientific management or efficiency movement, the idea that tests could prove or measure student learning grew in use at this time. Cubberly (1916) recommended using student performance on tests in conjunction with teacher efficiency rankings to determine the effectiveness of a teacher. Bobbit (1918) suggested the use of assessments to determine areas of strength and weakness for both teachers and students. In 1929, Wetzel suggested the idea that

teachers should be evaluated based upon measures of student learning and achievement. “No plan for evaluating any teacher’s work will be satisfactory to the teacher if it disregards the evidence of reliable pupil ratings” (p. 120). Wetzel recommended the use of aptitude tests for high school students to determine what students know and what they have learned. This concept of using tests and student performance is prevalent in today’s evaluation models.

After World War II, the use of observations as a part of teacher evaluations continued to grow (Marzano et al., 2011). The clinical supervision model based on the work of Cogan (1973) and Goldhammer (1969) became popular in the 1960s. “The process [clinical supervision] involved a purposeful, symbiotic relationship between practitioner and resident, where observation and discussion drove both parties to higher levels of growth and effectiveness” (Marzano et al., 2011, p. 18). The purpose for using the clinical supervision model in the classroom was to improve a teacher’s instruction (Cogan, 1973). The clinical supervision model focused on using observations as a formative assessment to provide feedback to teachers to improve their practice. There are five steps to the clinical supervision model: pre-observation and conference, observation, analysis, supervision conference, and a post conference analysis (Goldhammer, 1969, Pajak, 2003). Cogan (1973) added three more steps to the process, which were establish a relationship between teacher and supervisor, planning a strategy for the observation, and planning a strategy for the supervision conference. Cogan (1973) focused heavily on developing relationships and a collaborative focus. Goldhammer (1969) stressed the importance of using the method to develop reflection and self-assessment practices in teachers. With clinical supervision, the focus would be to improve instruction through observation, reflection, and dialogue. Many of the components of the clinical supervision model are evident in current coaching practices in schools.

The use of observations continued to be incorporated as a part of teacher evaluations into the 1970s and 1980s, but the Hunter Model was used to help guide observations. The Hunter Model was a template for effective lesson plans. Successful lesson plans, according to this model, had seven parts including (a) an anticipatory set to assess prior knowledge and focus student attention, (b) a statement of the objectives and purpose for learning, (c) the dissemination of information, (d) modeling of the skill or objective by the teacher, (e) a check for understanding through questions or other formative assessment techniques, (f) guided practice, and (g) independent student practice (Marzano et al., 2011). Almost serving as a checklist for observations, teachers were evaluated on the different components of the Hunter lesson plan model.

With the publication of *A Nation at Risk* by the National Commission on Excellence in Education in 1983, teachers and teacher effectiveness became the focus of school improvement. The report pointed to the dismal state of the education system and the need for highly qualified teachers to help improve learning for students (Shinkfield & Shufflebeam, 1995). Recommendations from the report regarding teachers and teacher evaluations were

- Set high standards for teacher performance.
- Improve teacher preparation programs.
- Offer salary and performance incentives based upon teacher evaluations.
- Create teacher career ladder systems to identify and distinguish effective teachers.
- Allow master teachers to be involved in the supervision and evaluation of new or novice teachers (National Commission on Excellence in Education, 1983).

The report garnered attention for the effective schools movement and led to increased research into what makes a successful teacher effective.

In 1984, soon after the *A Nation at Risk Report*, the RAND Corporation released the results of a study on teacher evaluation systems. According to the report, the purpose of teacher evaluations should be for teacher and school improvement, human resource decisions, and school status decisions. In studying the evaluation process of four different school systems, the authors made recommendations for successful evaluation systems. These included:

- The goals of the evaluation system should match the goals of the school system.
- The evaluation system should not be highly prescriptive.
- School systems should provide resources for evaluations such as time and personnel.
- Schools systems should assess the effectiveness of the evaluation systems and make changes as needed.
- Provide training and feedback for evaluators.
- Involve expert or master teachers in the supervision or evaluation process.
- Use standards to hold teachers accountable to their instructional decisions in the classroom (Wise, Darling-Hammond, McLaughlin, & Bernstein, 1984).

The *Nation at Risk Report* and the RAND report highlighted the need for effective teachers and reliable evaluation measures to determine teacher success.

Observations continued to be the main method for teacher evaluations into the 1990s. In 1996, Danielson published *Enhancing Professional Practice: A Framework for Teaching*, which provided a rubric for teacher evaluations that attempted to capture the complexity of teaching. The domains of the rubric were based on research of documented practices that improve student

learning. The framework evaluated teachers in the domains of planning and preparation, classroom environment, instruction, and professional responsibilities (Danielson, 2007). Similar rubrics and frameworks grew in their use at this time. The concepts represented in the Danielson rubric are replicated in many of the teacher evaluation rubrics in use today.

Currently, there is a heavy focus on accountability using student achievement to measure teacher effectiveness. Teacher accountability for student learning is a large focus of current evaluation models as a result of government regulations such as 2001's No Child Left Behind Act (2002) and 2015's Every Student Succeeds Act (2015). Many of the evaluation models in use today combine teacher observations using a rubric such as the *Framework for Teaching* (Danielson, 2007) with student achievement or gain data. Most of the student gain data is based upon student performance on standardized tests (Marzano et al., 2011). Value-added equations, such as TVAAS created by Sanders and McLean in Tennessee in 1980, assign a numerical score to represent how much impact a teacher has had on student learning (Shinkfield & Shufflebeam, 1995). These value-added scores are used as a part of teacher evaluations. The 2009 Race to the Top government program calls for teacher effectiveness to be determined by measures of student growth and observation based measures (United States Department of Education, 2016).

Tennessee Evaluation History

Tennessee followed the same developments and trends witnessed throughout the country in terms of teacher evaluations. In response to the recommendations from the *Nation at Risk Report* (1983), Tennessee implemented a career ladder for teachers in 1984 (Furtwengler, 1985). With the career ladder program, teachers could fall into one of five tiers: probationary, apprentice, career level I, career level II, and career level III. Teachers were placed into the

different tiers based upon experience, evaluation data, and recommendations. For example, a first year teacher was placed in the probationary tier. After one year, if the teacher had successful evaluations and recommendations from administrators, he or she could be moved into the apprentice tier. Evaluations from the career ladder program examined six different areas for teacher effectiveness. These areas were instructional preparation, instructional strategies, use of evaluation to improve instruction, classroom management, leadership role, and communication. Observations were the main data source for evaluations, but portfolios, interviews, and surveys could be used.

In 1997, Tennessee developed and adopted the *Framework for Evaluation and Professional Growth Model* (FEPG) for teacher evaluations. It was not implemented in schools until 2000. With FEPG, observations were used to evaluate teachers in the domains of planning, instructional strategies, classroom environment, assessment, professional growth, and communication. The number of observations that were required for a teacher was dependent on experience and performance. Teachers were either classified as apprentice or professional teachers. An apprentice teacher was a new teacher and could be moved to a professional with three years of teaching and successful evaluations. In addition to observations, apprentice teachers had to complete a comprehensive assessment, which involved a self-assessment, reflection, teacher information, and a growth plan. FEPG was discontinued and replaced with the current teacher evaluation model (Wright, 2012).

Tennessee Teacher Evaluation Models

In 2012, The Tennessee Teacher Advisory Committee reviewed different teacher evaluation models and rubrics, and chose the TEAM model as the official state evaluation model

(Tennessee Department of Education, 2012). The Teacher Advisory Committee also approved alternative evaluation models including the TIGER model. The TEAM model is used by a majority of the school systems in Tennessee, but 12 school systems have opted to use the TIGER model (Tennessee Department of Education, 2012). Both of these models represent the current trend in teacher evaluation models of combining measures of student growth, also known as value-added, with standards-based observations (Tennessee Department of Education, 2016b; Teacher Instructional Growth for Effectiveness and Results, 2016). In order to be an approved Tennessee teacher evaluation model, the evaluation model must include some observation instrument to measure the four areas of planning, environment, professionalism, and instruction. The approved models must also incorporate conferences, observations, and a review of past evaluations (Tennessee State Board of Education, 2015).

The TEAM model is the official Tennessee teacher evaluation model. The TEAM model uses the TEAM rubric to identify teacher performance in the four domains of planning, environment, professionalism, and instruction. The TEAM rubric is based on the ideas and principles of the National Institute for Excellence in Teaching evaluation model (Tennessee Department of Education, 2012). According to the Tennessee Department of Education (2012), the TEAM rubric "outlines clear expectations of high-quality instruction, planning, classroom environment and professionalism" (p. 9). When an evaluator visits a classroom to observe, a specific domain will be chosen as the focus of that observation. Half of the domain observations must be unannounced. The number of required observations is based upon the specific teacher's license and previous evaluation scores. A teacher's final score is a simple average of all of the indicator scores of all of the observations throughout the year. Based on evaluator ratings on the

TEAM rubric, the evaluator and teacher identify an area to focus on and an area in which the teacher is successful (Tennessee Department of Education, 2016b).

The TEAM model incorporates coaching conversations or conferences with teachers. At the beginning of the school year, principals are expected to meet with teachers to discuss the teacher's previous performance. Teachers who earned a score of a one on the previous year's evaluation will have a focused conversation about observations and about supports that will be provided in order to help them improve. The coaching sessions should incorporate data from both formative and summative assessments and should enable the teacher to look more closely at patterns in his or her own teaching. The overall goal is to improve teacher performance and to increase student learning (Tennessee Department of Education, 2016b). TEAM coaches are available to schools where there is a large difference between observation scores and value-added scores. These coaches work with evaluators to help them with observations, facilitating effective coaching conversations, providing useful feedback, and providing appropriate professional development opportunities (Tennessee Department of Education, 2015).

The TIGER model for teacher evaluations is structured somewhat differently. The goal of the TIGER model is to, "promote teacher growth and ensure all teachers provide quality instruction" (Teacher Instructional Growth for Effectiveness and Results, 2016, para. 1). Guiding principles of the TIGER model include:

- teacher evaluations are based upon high quality standards of performance;
- focus on teacher growth and development;
- includes both formative and summative assessments;
- provides differentiated support based on teacher performance;

- provides coaching for Stage I teachers;
- provides leadership opportunities for Stage III teachers;
- supports the use of professional learning communities;
- incorporates both qualitative and quantitative data (Teacher Instructional Growth for Effectiveness and Results, 2016).

The TIGER model also uses the TEAM rubric to evaluate teachers during classroom observations. Unlike the TEAM model, the observations are not focused on specific domains for each observation. The evaluator considers all domains during an observation. The TIGER method also uses walkthroughs, or short classroom visits, in order to document daily activity in the classrooms and to provide formative assessment data for teachers. These walkthroughs are used to provide feedback to teachers to help improve performance. The final score for a teacher with the TIGER model is not an average of all of the observation scores. The principal or administrator considers all observation data, walkthrough data, and teacher artifacts to indicate where a teacher is performing on the TEAM rubric. The use of teacher artifacts is unique to the TIGER model. If an observer did not see a specific indicator during walkthroughs or observations, a teacher can provide his or her own evidence or artifact to prove their success with that indicator. Possible artifacts include lesson plans, sample assignments and assessments, examples of communication with parents and guardians, and a professional development plan (Teacher Instructional Growth for Effectiveness and Results, 2016). Teachers may submit evidence at any time throughout the year.

Based on final evaluation scores, TIGER teachers will be placed in one of three stages. Stage I teachers are new teachers or teachers in need of improvement. Stage II teachers are teachers that are rated as effective, and stage III teachers are highly effective. Like the TEAM model, the

number and frequency of observations is based on the performance or stage of the teacher (Teacher Instructional Growth for Effectiveness and Results, 2016). Stage 1 teachers are required to have a coaching component throughout the year. The teacher and coach will review teacher performance information and will work together to create a growth plan. The coaching conversations and walkthroughs are required to happen at least twice a semester. Stage III teachers can have leadership opportunities.

Both the TEAM and TIGER models use the same method for scoring teachers in terms of effectiveness. As mentioned earlier, 35% of a teacher's evaluation comes from TVAAS, 15% comes from a different measure of student growth, and the remaining 50% is from observations for teachers in subject areas that are tested. When the scores are computed, teachers are given a numerical score from one to five. A teacher who earns a five has performed significantly above expectations and has shown much growth for his or her students. A teacher who earns a four performed above expectations, but did not show as strong of a growth for his or her students. A score of a three means that a teacher met expectations and had the expected amount of growth for his or her students. A score of a two means that teacher performed below expectations and did not have as much student growth as expected. A score of a one signifies a teacher who was significantly below expectations with little to no growth for students (Tennessee Department of Education, 2016b).

Current research regarding the TEAM and TIGER evaluation models is sparse. Much of the available information focuses on teacher perceptions of the two models as measured by survey responses. Since implementation of the First to the Top Act, the Tennessee Department of Education has surveyed teachers to gain feedback and insight into their feelings regarding the evaluation process through the First to the Top Survey (Ehlert et al., 2013; Pepper et al., 2015).

The survey was distributed to all educators in Tennessee's Educator Information System. Participation in the survey was voluntary. The survey results offered valuable information about the two models. When asked in 2014 if teachers were satisfied with the evaluation process, 45% of TEAM teachers agreed or strongly agreed while 63% of TIGER teachers agreed or strongly agreed (Pepper et al., 2015). Results were similar with an earlier First to the Top Survey (2013) where 68% of TIGER teachers agreed or strongly agreed and 48% of TEAM teachers agreed or strongly agreed (Ehlert et al., 2013). When asked about follow up meetings after observations, 48% of TEAM teachers had one or more follow up sessions and 62% of TIGER teachers had one or more follow up sessions (Ehlert et al., 2013). When asked the same question in 2015, 50% of TEAM teachers indicated they had at least one follow up meeting compared to 49% of TIGER teachers (Pepper et al., 2015). Teachers were also asked to self-report regarding how much they thought they improved as teachers during the 2013-2014 school year. Twenty-eight and a half percent of TEAM teachers felt that they improved significantly compared to 29.7% of TIGER teachers (Pepper et al., 2015). "Perceptions about the quality, benefits, and value from teacher evaluations from teachers in districts using TEAM were consistently less favorable than the perceptions of teachers in districts using one of the alternative models, TEM, TIGER or COACH" (Ehlert et al., 2013, p. x).

Tennessee also invited teachers to participate in the Tennessee Teaching, Empowering, Leading, and Learning Survey (TELL). The goal of the survey was to discover teacher feelings regarding teaching conditions. Like the First to the Top Survey, all Tennessee educators were invited to take part in the survey. The survey focused on nine different work conditions: time, facilities and resources, community support, managing student conduct, teacher leadership, school leadership, professional development, instructional practices and support and new teacher

support. In each of the nine categories, teachers using the TIGER model had a higher percentage of agreement than teachers in the TEAM model. Teachers were also asked if they felt that their performance was measured objectively by evaluators. Eighty-three percent of TEAM teachers agreed as compared to 89% of TIGER teachers. Teachers were also asked if they felt that they were held to high standards. Ninety-six percent of TEAM teachers agreed and 97% of TIGER teachers agreed (New Teacher Center, 2013).

Most of the available information regarding both models has come from the Tennessee Department of Education or affiliated groups. In the year one implementation report (2012), TVAAS scores across the different models were shared (see Table 1). Results indicate that TIGER teachers had higher percentages of teachers earning higher growth scores (Tennessee Department of Education, 2012, p. 33). Fifty-nine percent of TIGER teachers earned a score of a four or a five as compared to 50.8% of TEAM teachers.

Table 1

Individual TVAAS Composite Score by Evaluation Model

TVAAS	**1**	**2**	**3**	**4**	**5**
TEAM	16.4%	8.1%	24.7%	11.9%	38.9%
TIGER	15.2%	5.8%	20.3%	11.5%	47.5%

The year one implementation report also shared the distribution of observation scores across the different evaluation models (see Table 2). There were more teachers with lower observation scores using the TEAM model when compared with the TIGER model (Tennessee

Department of Education, 2012, p. 33). Twenty-three percent of TEAM teachers received a score of a one or two on observations compared to 1% of TIGER teachers.

Table 2

Individual Observation Scores by Evaluation Model

Observation	**1**	**2**	**3**	**4**	**5**
TEAM	0.2%	23.0%	22.7%	52.8%	22.0%
TIGER	0.1%	1.0%	14.6%	67.4%	17.0%

Most of the available information regarding evaluation models has been provided by the Tennessee Department of Education or its affiliates. A majority of this information has come from survey data. There is very little published research focusing on these evaluation models. There are a few studies that have examined at the TEAM model, but not the TIGER model. This could be attributed to the fact that the TEAM model is the official Tennessee state evaluation model. Davis (2014) specifically examined the relationship between observations scores in the TEAM model and student achievement as measured by a teacher's TVAAS score. While limited by the fact that the study was only performed in one school system, Davis found a weak relationship between TVAAS and TEAM teacher ratings (Davis, 2014). The Tennessee Department of Education (2015), in the year three implementation report, claimed that "over time we have also seen a closer correlation between observations and student growth measures at levels 4 and 5" (p. 13). This increase in correlation has been linked to increased training for evaluators and the use of TEAM coaches (Tennessee Department of Education, 2015). Mason (2015) found a relationship between TEAM observation scores and student achievement as

measured by TVAAS in English II and III classrooms. Mason only focused on the thinking and problem solving indicators on the TEAM observation rubric and discovered that the thinking indicator had a strong, positive, relationship to student achievement as measured by performance on the End of Course exam (Mason, 2015).

Teacher and principal perceptions of the TEAM model have also been studied. A 2013 study by Bryant investigated principals' perceptions of the TEAM model and the impact of the model on professional growth through the use of a survey. Results indicated that principals had positive beliefs about the TEAM model's ability to support professional growth and development. The principals reported that the model provided a standard process for evaluations and also helped improve teacher instruction through feedback. The biggest hindrance to this model as noted by participants was the time required to effectively implement the TEAM evaluation model (Bryant, 2013). Bogart (2013) explored teacher perceptions of the TEAM model also through a survey. Teachers were surveyed after implementation of the TEAM model. Teachers indicated that implementing the TEAM model impacted how they planned lessons. Survey results indicated that teachers were focusing more on higher order thinking skills and that teachers considered the TEAM rubric more as they planned lessons. Teacher perceptions about the evaluation process did not change significantly with the implementation of the TEAM model (Bogart, 2013).

Value-Added

In addition to observation scores, teachers are evaluated based upon student growth. This student growth is often known as value-added. "Value added refers to a particular approach to measuring test score gains. What distinguishes value-added measures from simpler test score

growth measures is that they are adjusted for between-classroom differences in student characteristics" (Ferguson & Danielson, 2014, p. 101). This topic has been the subject of many recent studies in education. Multiple researchers have completed reviews of the research to identify strengths and weaknesses with this measure. Papay (2012) identified the following strengths to value-added models: they measure growth, they are objective, they are easy to compute, and they focus on outputs. With these strengths, researchers have also identified weaknesses. Value-added scores are often inconsistent over time (Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2012; Hallinger, Heck, & Murphy, 2014). Growth scores are subject to bias because students are often not randomly assigned to teachers. Frequently, high or low performing students are assigned to specific teachers instead of being randomly placed in classrooms **(**Darling-Hammond et al, 2012**;** Hallinger et al., 2014**;** Papay, 2012). Value-added formulas are not able to account for all of the factors that improve student learning (Amrein-Beardsley, 2008; Darling-Hammond et al, 2012; Hallinger et al., 2014; Papay, 2012). Because value-added is simply only given with a numerical measure, it is difficult for teachers to receive effective feedback showing them specific areas for improvement (Papay, 2012). Darling-Hammond et al. (2012) also found that value-added measures do not often relate to evaluator scores. There is also a concern that the assessments used to compute value-added scores do not accurately represent the curriculum standards making the value-added measure invalid (Amrein-Beardsley, 2008).

Many researchers have investigated how well teacher observations are related to value added scores. A 2014 study by Harris, Ingle, and Rutledge of principal evaluations found a weak relationship between principal evaluations and value-added scores. It was hypothesized that this difference could be due to different principals having different priorities and traits valued

(Harris, Ingle, & Rutledge, 2014). Perhaps this is why Goldring et al. (2015) found that many principals do not believe that value-added measures are valid. A 2011 study by Kane et al. found the opposite. The study examined the relationship between value-added scores and observation scores, and found a positive correlation between the two (Kane et al., 2010). Jacob and Lefgren (2008) found that value-added scores more closely correlated with a principal's personal evaluation of a teacher rather than the teacher's observation scores.

Gill, Shoji, Coen, and Place (2016) examined teacher evaluation and student achievement data from a prior MET study investigating five different observation instruments. These five observation instruments were the Classroom Assessment Scoring System, the Framework for Teaching, The Protocol for Language Arts Teaching Observation, the Mathematical Quality of Instruction, and The UTeach Observational protocol. The researchers examined how well the different dimensions on these rubrics correlated with student growth or value-added scores. The researchers found a moderate, but statistically significant relationship among all of the teaching dimensions on these instruments and value-added scores.

A 2014 study by Staiger and Kane investigated the ability of value-added over many years to predict a teacher's career effectiveness. The researchers acknowledged that value-added scores often fluctuate from year to year for individual teachers, and wanted to investigate whether value-added was an accurate measure for long term success of teachers. This study involved math and language arts teachers from three school districts who had between six and nine years of value-added data. The researchers examined the average value added scores for these teachers over the long term. They called this the year-to-career calculation. Staiger and Kane (2014) found that, "accumulated value-added estimates averaged over a teacher's career to date, are better predictors of future value added and are considerable more stable than single-year

value-added estimates" (p. 160). While there may be fluctuation from year to year, the researchers found a correlation between value-added scores and long term teacher performance.

Because there have been some questions regarding the use of value-added measures as a part of teacher evaluations, Polikoff (2014) studied the effectiveness of incorporating different measures as a part of teacher evaluations. Polikoff investigated optimal weights of three different components to teacher evaluations. These three components were value-added scores, teacher observation scores, and student survey data. Polikoff examined the different weights for these three measures in six different states and concluded that the optimal weights for these three measures are as follows. Value-added measures should count for 75%, observation scores should represent 11%, and survey results should make up 14% of the total. Polikoff argues that this ideal weighting leads to greater stability in the overall measure of teacher effectiveness. Polikoff did acknowledge that this weighting system may not be appropriate for every state due to differences in the ability of each state's test to measure what is taught. With this acknowledgement, Polikoff (2014) still believes that, "the stability of teacher performance ratings is higher when the composite includes a greater weight on FfT [Framework for Teaching observation scores] and Tripod [student survey] scores, since those individual components have higher reliability than VAM [value-added measures] scores do" (p. 299). Polikoff's conclusions signal the importance of using multiple measures to score teacher effectiveness. Value-added is not enough to measure the impact of a teacher. Current Tennessee evaluation models factor in multiple variables into a teacher's evaluation score.

Tennessee Value Added Assessment System

History of TVAAS. TVAAS originated in the 1980s from the work of Sanders and McLean (Kupermintz, 2003; SCORE, 2014). The system was created with the purpose of measuring the impact of a teacher on student learning. With the passing of the Educational Improvement Act in 1992, TVAAS was adopted by the State of Tennessee (Baker & Xu, 1995; SCORE, 2014; Smith, 2004). The goal with implementing TVAAS was to increase accountability. In the Master Plan for Tennessee Schools in 1993, one of the listed goals for the education system was accountability. The accountability focus meant that there should be a focus on educational results as well as a system to assess the results. Within this accountability goal, creating and maintaining a value-added system was given as a strategy to increase accountability (Tennessee State Board of Education, 1992). In 1993, district level TVAAS reporting began followed by school level reporting in 1994. Teacher level TVAAS reporting did not being until 1996 (SAS Institute, 2015). With the passage of the First to the Top Act in Tennessee in 2010, individual teacher TVAAS ratings became a part of a teacher evaluation scores (SCORE, 2014). Currently, a teacher's TVAAS score counts for 35% of a teacher's evaluation score (Tennessee Department of Education, 2016b). In 2013, the state of Tennessee amended the teacher evaluation policy to allow a teacher to include his or her TVAAS score as their additional 15% measure in evaluations if the score would help improve his or her overall evaluation score (SCORE, 2014).

TVAAS description. With the passing the Educational Improvement Act, TVAAS has been an accountability tool in the state of Tennessee since 1992. According to the Tennessee Code Annotated (2016):

(a) “Value added assessment” means:

(1) A statistical system for educational outcome assessment that uses measures of student learning to enable the estimation of teacher, school and school district statistical distributions; and

(2) The statistical system will use available and appropriate data as input to account for differences in prior student attainment, such that the impact that the teacher, school, and school district have on the educational progress of students may be estimated on a student attainment constant basis. The impact that a teacher, school and school district has on the progress, or the lack of progress, in educational advancement or learning of a student is referred to hereafter as the “effect” of the teacher, school, or school district on the educational progress of students.

(b) The statistical system shall have the capability of providing mixed model methodologies that provide for the best linear unbiased prediction for the teacher, school and school district effects on the educational progress of students. It must have the capability of adequately providing these estimates for the traditional classroom of one (1) teacher teaching multiple subjects to the same group of students, as well as team taught groups of students or other teaching situations as appropriate.

(c) The metrics chose to measure student learning must be linear scales covering the total range of topics covered in the approved curriculum to minimize ceiling and floor effects. These metrics should have strong relationship to the core curriculum for the applicable grade level and subject. (Tennessee Code Annotated, 2016, §49-1-603)

For the 2014-2015 school year, the measures of student learning used were the Tennessee Comprehensive Assessment Program (TCAP) for grades three through eight and the End of Course Exam (EOC) for high school students. In order for an assessment to be used in the TVAAS calculation, three requirements must be met: the test needs to be able to show growth for both high and low performing students, the test must address the academic content standards, and the measurement scales for the test must be reliable from year to year (SAS Institute, 2015).

To determine growth, TVAAS uses a comparison of a student's current achievement with his or her past achievement on the TCAP or EOC assessment (SAS Institute, 2015). TVAAS predicts a student's performance based on prior achievement as well as the average growth of similarly performing students in the state (SCORE, 2014). A mixed model equation containing the above factors is used to determine a student's growth (Sanders & Horn, 1994). A Multivariate Response Model is used for tests like the TCAP, which are given consecutively. A Univariate Response Model is used for tests like the EOC, which is given to students in multiple grades (SAS Institute, 2015). To calculate a teacher's effect, a layered model is used. The model considers the effects of the current teacher as well as previous teachers, all available testing data for a student, and the percentage of time a student has spent with the teacher (SAS Institute, 2015). A TVAAS score of a zero indicates that a student has grown as expected. Any number greater than zero indicates growth that was greater than predicted, while a negative score indicates growth that was less than expected.

According to the Tennessee Department of Education (2016b), there are many reasons and purposes for using TVAAS in schools. TVAAS can help to monitor the progress and achievement levels for all students. TVAAS can also provide teachers and administrators with data in order to improve instruction and provide direction for professional development needs.

TVAAS also helps to ensure that all students will be challenged to grow and learn as teachers attempt to have learning gains for all groups of students. TVAAS can also help teachers and administrators to identify what is working in classrooms. It is also an opportunity to celebrate and acknowledge growth and achievement.

Perceived disadvantages of TVAAS. The use of value-added measures has come under scrutiny. There have been concerns over the validity of the measure for different reasons. There is concern that the test is not valid because there is so much variability in growth scores for the same teacher from year to year (Baker & Xu, 1995; SCORE, 2014). Another concern is that there are so many factors that affect student achievement that are not included in the mixed model equation, and so the value-added model cannot validly measure a teacher's impact (Baker & Xu, 2015; Kupermintz, 2013; SCORE, 2014). Because the model does not take into account socioeconomic status or student demographic information, there are concerns that the formula is not valid. These factors have been shown to impact student learning (Ballou, Sanders, & Wright, 2004). Others believe that because students are not randomly assigned to teachers, the model is not reliable. In many schools, successful teachers have the higher performing students, while new or underperforming teachers receive the lower performing students. Random assignment does not often occur (Kupermintz, 2003; Sanders & Horn, 1994). In addition to students not being randomly assigned, teachers with less data (from having less students), would have less accurate assessments of value-added. More data increases the accuracy of the value-added score (Kupermintz, 2003). A final area of concern relates to the assessments used to measure achievement. There are concerns that the TCAP test and the EOC exams do not accurately measure knowledge of the content standards, so the data from these assessments would not show how effective a teacher was in teaching the standards. Also, there is concern that these

assessments do not ask enough challenging questions to show growth for the higher performing students, thus making it harder for a teacher to show growth with high performing students (SCORE, 2014).

Perceived advantages to TVAAS. There are many concerns with the value-added model, but there are also many reasons for why this approach has become so popular. The complex formula can accommodate test data that measure achievement on different scales. This means that past student data can be used even if a student has a different testing history from other students. Measurement error in the calculation is minimized due to the inclusion of all prior testing data from different subjects and grade levels. Another benefit is that the layered model that is used to figure out a teacher's value-added, which can accommodate for team teaching situations and other shared teaching situations (SAS Institute, 2015).

TVAAS research. TVAAS has been used in many research studies as a means to measure student achievement or teacher effectiveness. There is very little academic research regarding the accuracy or use of TVAAS. Among the available research, most has been published by Sanders, the creator of TVAAS, and thus calls the results into question. One study performed by Ballou et al. (2004) was conducted to address the issue that TVAAS is not valid because student socioeconomic status and demographic information are not included in the formula. The creators have argued that the students serve as their own control, and so these measures do not need to be included. In this study, the researchers added demographic and socioeconomic status information to the formula and compared the results to when these factors were not included. The researchers concluded from the results that the inclusion of these factors in the TVAAS calculation had little impact on the scores, and believed they were justified in not having those factors in the formula.

A 1997 study by Wright, Horn, and Sanders examined teacher effectiveness, as measured by TVAAS, when controlling for classroom heterogeneity, student achievement level, and class size. The researchers examined TCAP scores for third through fifth graders for the purpose of this study. Results of the study indicated that class size did not have any significant impact neither did the heterogeneity of the classroom. In terms of the composition of the classroom, results indicated that successful teachers can help students at all achievement levels to learn regardless of the heterogeneity of the class. There was a significant impact with achievement levels. The researchers found that the highest performing students had smaller gains than the lower performing students.

A 1994 study by Sanders et al. was performed to examine drops in student achievement levels, as measured by TVAAS, from grade to grade. Specifically, the researchers hypothesized that drops in scores were caused by transition to new buildings. Using Tennessee's student TCAP achievement data from one year, the researchers found that there was a drop in achievement when students transitioned to a new building (i.e. from middle school to high school or elementary school to middle school). The drop did not appear to happen when a new student transferred in from another system. Results from this study indicate that teachers who teach a grade level that is the youngest in the building could possibly expect lower TVAAS growth due difficulties that students may have with the transition to a new building.

Sanders and Rivers (1996) were interested in studying teacher effects on student achievement, specifically if teacher effects impacted future student achievement gains. This study used TCAP and TVAAS data from math teachers and students in grades three through five in two large Tennessee school systems. Similar to the 1994 study by Sanders et al., the data suggested that effective teachers can help all levels of students, especially lower performing

students. The researchers also found that teacher effects can raise achievement, and the effects of a teacher do impact future learning. “An effective teacher receiving students from a relatively ineffective teacher can facilitate excellent academic gain for his/her students during the school year. Yet these analyses suggest that the residual effects of relatively ineffective teachers from prior years can be measured in subsequent student achievement scores” (Sanders & Rivers, 1996, p. 4).

Standards Based Observations

Both the TEAM and the TIGER teacher evaluation models incorporate observations using use the same TEAM rubric to rate teachers during classroom observations. The rubric bases teacher performance in four domains: instruction, planning, environment, and professionalism (Tennessee Department of Education, 2016b). Standards based observations judge the quality of teaching against set standards of effectiveness usually with some instrument such as a rubric (Heneman, Milanowski, Kimball, & Odden, 2006). Papay (2012) reviewed the available research regarding the use of observations in teacher evaluations and was able to identify strengths and weaknesses with this evaluation method. Some of the strengths of standards-based observations include the ability to describe teacher performance levels through rubrics, and the evaluation is based on evidence of effective teaching. Weaknesses of the observation method include the amount of time required to complete the observation, the possibility of evaluator bias, and the lack of reliability in scores between different evaluators (Papay, 2012). In a study of teacher perceptions regarding the evaluation process, a majority of teacher participants indicated that they felt that observations provided them with useful feedback in order to improve teaching practices, and teachers felt that observations accurately measured their performance (Jiang, Sporte, & Luppescu, 2015). A large portion of the research on

standards based observations has focused on how well observation measures relate to student achievement and the ability of observers to accurately evaluate teachers. In both areas, the research presents a mixed view.

In 2009, Wesiberg, Sexton, Mulhern, and Keeling surveyed teachers and administrators in school systems in four states to get a better understanding of how teachers are scored or ranked based upon observation data. The researchers found that in school systems that only identify teachers as satisfactory and unsatisfactory, 99% of teachers were labeled as satisfactory. In school systems that have more ranking categories, 94% of teachers were in the top two categories. When looking at the scoring of new teachers, 66% of new teachers were evaluated as satisfactory. This led the researchers to conclude that observations alone are not effective at identifying successful teachers.

Much of the research on standards-based observations has centered on the relationship between teacher evaluation scores from observations and student growth or achievement. Student achievement is often measured using growth or value-added measures calculated from student performance on standardized tests (Gallagher, 2004; Garret & Steinberg, 2015; Grossman et al., 2014; Kimball, White, Milanowski, & Borman, 2004). Results of these studies are contradictory as some found a positive correlation between observation scores and student achievement and some found no correlation between observations and student growth. Gallagher (2004) studied elementary teachers' performance ratings from observations to see how well scores correlated with value-added measures. These teachers all worked in the same school system, and the observations were performed with a rubric that contained content specific domains. The evaluation process involved conferences with the evaluator before and after the observation. Results of the study indicated a strong relationship between literacy evaluation scores and value-

added. Gallagher (2004) found that student performance on standardized assessments averaged 14 points higher with a one-point increase in the teacher evaluation. No relationship was found between the evaluations scores in mathematics and language arts and student achievement. This study was unique because it involved a content-specific rubric. Hill and Grossman (2013) argued that teacher observation rubrics or instruments should focus on the specific subject areas because there are special skills and teaching practices that are more effective in the different subject areas.

Grossman et al. (2014) found that the type of assessment matters when looking for a relationship between value-added and observation scores. Teachers in this study were evaluated with the PLATO rubric, an observation instrument specific for language arts instructors. The students were given the state standardized test and the SAT-9 test. When comparing results, it was discovered that the observation results had a stronger correlation with the value-added measures calculated from the SAT-9 test over the state standardized test. Researchers determined the stronger relationship existed with the SAT-9 test because this assessment more closely measures skills that are valued in the PLATO rubric (Grossman et al., 2014). Results of this study indicate that the type of assessment given is important to the strength of the relationship between teacher performance and value-added growth measures.

Two similar studies identified a positive relationship with teacher observation scores and value-added scores. Garret and Steinberg (2015) investigated the relationship between teacher observations and student achievement. Specifically, the researchers we interested in understanding if effective teachers could be successfully identified using the *Framework for Teaching* rubric. They used student achievement to identify effective teachers. The researchers found that in both language arts and math, higher performing teachers as identified by the rubric

had higher performing students. In language arts, the difference between a basic and proficient teacher translated into a 0.11 standard deviation difference in student achievement. In math, the difference between a basic and proficient teacher, translated into a 0.14 to 0.16 standard deviation in student achievement. The results were statistically significant. The researchers acknowledged that there was some bias to these results due to the fact that students were not often randomly assigned to teachers (Garret & Steinberg, 2015). A 2004 study by Kimball et al. also used an evaluation system modeled after the *Framework for Teaching* rubric to see if the scores correlated to student achievement. The researchers found a positive correlation, but this relationship was not as strong when certain student factors such as prior achievement were controlled. The teachers in this study were observed as part of their evaluation, but there were no consequences for the teachers for low scores. Both of these studies found a correlation, but both also admitted to limitations with each study.

Similar studies performed have also indicated weak or no relationship between evaluations and student achievement. Davis (2014) examined the relationship of the observations scores in the TEAM model and their relationship with student achievement as measured by a teacher's TVAAS score. Though limited by the fact that the study was only performed in one school system, Davis found a weak relationship between TVAAS and TEAM teacher ratings (Davis, 2014). Kimball and Milanowski (2009) found a weak relationship between observations ratings and student achievement. The researchers in this study attribute the inconsistency to differences in evaluators (Kimball & Milanowski, 2009). Hallinger et al. (2014) researched the available studies regarding teacher evaluations and found a similar weak relationship between evaluations and student learning as measured by value-added measures.

Because some of the available research has shown a weak relationship between observations and value-added, some researchers have begun to investigate reasons why this is the case. Multiple studies have focused on the evaluators themselves and how they factor into observation scores. Jacob and Lefgren (2008) investigated principals' ability to recognize high and low performing teachers from teacher observations. Results of the study indicated that the principals were able to identify the highest and lowest performing teachers from observations, but they were unable to differentiate between middle performing teachers (Jacob & Lefgren, 2008). The principals in this study were evaluating their own teachers. A 2011 study also investigate how well evaluators were able to identify the performance of teachers through evaluations, but this study involved evaluators who did not know or have any prior experience with the teachers. The researchers found that the evaluators were not consistently able to identify the high and low performing teachers (Strong, Gargani, & Hacifazlioglu, 2011). The accuracy of teacher evaluations in rating teacher performance is important because as Goldring et al. discovered in their 2015 study, principals will often use observation scores and experiences to make decisions regarding employment rather than using value-added scores.

A 2014 study by Park, Chen, and Holtzman investigated whether a specific training program for observers and raters of teachers could reduce bias and increase the effectiveness of the raters. This study was undertaken to help discover ways to reduce rater bias. The researchers define rater bias as, "personal preferences, viewpoints, or interpretations of the instrument that are external to the scoring rubric" (p. 384). To attempt to reduce rater bias, the researchers investigated the impact of a rater training program. The evaluators in this program were mainly teachers or former teachers and education students. Participants were trained to use a specific performance rubric as well as ways to identify and eliminate bias in evaluations. Training was

given online. Raters were required to pass a certification test to be assigned as evaluators. After the initial certification, the evaluators were monitored and re-evaluated to ensure effectiveness. Park, Chen, and Holtzman were interested in the effectiveness of the training as well as influences on scoring. The raters evaluated teacher's recorded lessons. Each recording was evaluated by two raters. Of all of the scoring results, the researchers found that over 75% of the observations showed complete agreement or agreement within one point from both evaluators. The researchers also found that classroom settings, rater background and experience, rater perceptions, and attention to detail did not significantly influence the observation scores. This study has implications for the creation of successful teacher evaluation programs. With appropriate training, evaluators could be more successful with evaluations and could eliminate bias in their scoring. The researchers suggested that a successful teacher evaluation model should include in-depth training, frequent monitoring of evaluators, and feedback with assistance for ongoing development of evaluation skills.

In a further effort to study evaluators and observations, Bell et al. (2014) conducted a study to determine which areas in teacher observations were the easiest and most challenging to observe and evaluate. Teachers in grades six through eight were evaluated using video recordings of mathematics or language arts lessons. Observers used the Classroom Assessment Scoring System rubric, the Framework for Teaching Rubric, the Mathematical Quality of Instruction rubric, and the Protocol for Language Arts Teaching Observation to evaluate the teachers. The Mathematical Quality rubric and the Protocol for Language Arts are content specific classroom observation rubrics. Think aloud data and interviews with observers were used to answer the research questions. Based on the data, the researchers concluded that observation domains that required less inference were the easiest to measure such as classroom

interactions and classroom environment. Domains that required higher levels of inference were the most challenging areas for observers to evaluate such as instruction (Bell et al., 2014). Results of this study are important as they signify that not all areas of teacher evaluations can be measured objectively, and as a result more subjectivity may go into evaluation scores when higher levels of inference are required.

Joe, McClellan, and Holtzman, (2014) studied the impact of observation time and the number of domains being evaluated on the accuracy of teacher observation scores. There were two parts to this study. The first part investigated if length of time or the time period of the observation had any impact on the accuracy of the observation. The researchers discovered that when scoring teachers using the CLASS, FfT, or QST rubric, the scores from minutes 15 to 30 (segment 2) and minutes 30 to 45 (segment 3) had a strong relationship to the overall score of effectiveness for the teacher. These results suggest that observations do not need to occur throughout a whole lesson to accurately measure teacher success. The results also indicate that observing teachers for 30 minutes during minutes 15 to 45 of a lesson are enough to get an accurate picture of teacher effectiveness. The second part of the study examined the impact of only measuring some of the observation rubric domains on the accuracy of the observation. Observers were only asked to score teachers in a few of the rubric dimensions. The researchers found that there was higher inter-rater reliability when observers were not asked to score teachers based on all of the domains. The researchers suggested that this was due to the fact that limiting the number of domains reduced the cognitive load for observers. These results suggest that only evaluating specific domains for each observation may increase the accuracy of the observation. Observations using the TEAM model are set up in such a manner where each observation focuses on a specific domain.

Chapter Summary

Chapter 2 presented related literature for this study. Literature regarding teacher evaluations, teacher effectiveness, observations, and value-added measures were shared. Much of the available research has provided mixed results. Most current teacher evaluation systems incorporate value-added measures as well as observation measures, but research is inconclusive as to whether these measures accurately measure effective teaching or successfully represent the characteristics of successful teachers. Value-added measures are easy to compute and are objective, but are subject to concerns over accuracy and reliability. Observations allow for a greater understanding of a specific teacher's classroom practices, but are subject to concerns over the reliability of evaluators to accurately score observations without bias. Research indicates that effective teachers are successful at managing the classroom, differentiating content, and making lessons authentic.

CHAPTER 3

RESEARCH METHODOLOGY

This study was conducted to investigate the relationship between measures of teacher effectiveness within teacher evaluation models. The purpose of this study was to specifically compare the Tennessee Educator Acceleration Model (TEAM) and the Teacher Instructional Growth for Effectiveness and Results Model (TIGER) on measures of teacher effectiveness. Teachers' TVAAS growth score along with final observation scores from administrators, from either the TEAM or TIGER model, were used to measure teacher effectiveness in this study. The relationship between value-added growth scores and observation scores was also compared between the two evaluation models.

A quantitative approach was used to compare the relationship between measures of teacher effectiveness and teacher evaluations models. A quantitative approach is appropriate for numerical data, performance data, and closed-ended research questions (Creswell, 2002; Creswell, 2014). This chapter includes: the research design, the research questions and corresponding null hypotheses, instrumentation, population, data collection, data analysis, and a summary.

This quantitative study compares measures of teacher effectiveness between the TEAM and TIGER Tennessee teacher evaluation models using a non-experimental, causal comparative approach. According to Creswell (2014), a quantitative approach is used to investigate relationships between variables that can be measured numerically. A teacher's final observation score can range from one to five. An individual teacher's TVAAS growth score can also range

from one to five. A quantitative approach is also appropriate for closed ended research questions that are explanation oriented (Creswell, 2002; Creswell, 2014).

A non-experimental study is used to, "examine relationships between different phenomena without any direct manipulation of conditions that are experienced" (McMillan & Schumacher, 2010, p. 22). This study used pre-existing data from the 2014-2015 school year. School systems were not assigned to be TEAM or TIGER schools. Each school system previously decided on its own which evaluation model to use. This research was simply interested in looking at the existing data to determine if there was any significant difference in measures of teacher effectiveness between the two evaluation models.

A causal comparative non-experimental design was used for this study. A causal comparative approach is appropriate for this study because the research involved pre-existing intact groups without any manipulation of variables (Creswell, 2014). With causal comparative research, the independent variables are categorical, and in the case of this research, the independent variable was the teacher evaluation model (Schenker & Rumrill, 2004). The goal of causal comparative research is to investigate differences between groups.

Research Questions and Null Hypotheses

The following research questions and null hypotheses were identified for this study:

Research Question 1

Is there a significant difference in overall teacher effectiveness as measured by TVAAS between teachers who are evaluated with the TEAM model and teachers evaluated with the TIGER model?

H_{01}: There is no significant difference in overall teacher effectiveness as measured by TVAAS between teachers who are evaluated with the TEAM model and teachers evaluated with the TIGER model.

Research Question 2

Is there a significant difference in final principal observation scores between teachers who are evaluated with the TEAM model and teachers evaluated with the TIGER model?

H_{02}: There is no significant difference in final principal observation scores between teachers who are evaluated with the TEAM model and teachers evaluated with the TIGER model.

Research Question 3

Is there a significant relationship between teachers' value-added growth scores (TVAAS) and teachers' final observation scores for teachers who are evaluated with the TEAM model?

H_{03}: There is no significant relationship between teachers' value-added growth scores (TVAAS) and teachers' final observation scores for teachers who are evaluated with the TEAM model.

Research Question 4

Is there a significant relationship between teachers' value-added growth scores (TVAAS) and teachers' final observation scores for teachers who are evaluated with the TIGER model?

H_{04}: There is no significant relationship between teachers' value-added growth scores (TVAAS) and teachers' final observation scores for teachers who are evaluated with the TIGER model.

Instrumentation

In order to measure teacher effectiveness, individual teacher value-added growth scores (TVAAS) and individual final classroom observation scores from the teacher's principal with the TEAM rubric were used. These two measurements were chosen because they are the two largest factors that are included in scoring for overall teacher effectiveness in Tennessee teacher evaluations. According to the SAS Institute (2015), TVAAS measures are reliable because they are based on assessments that allow for the measurement of progress for both high and low achieving students, they are aligned to the academic standards, and the measurement scales are reliable from year to year. The use of an intra-year approach (multiple years) is also another reason given for the reliability of TVAAS growth scores (SAS Institute, 2015).

The TEAM observation rubric is used by administrators for both the TIGER and TEAM evaluation models to document the use of effective teaching practices within the classroom. For each indicator on the TEAM rubric, an administrator rates the teacher on a scale from one to five. Evaluators using this rubric are trained and assessed regarding their ability to successfully use the rubric for evaluations. In order to observe and evaluate teachers, administrators must be certified, and certification for using the rubric must be obtained yearly through either in-person training or an online recertification test (Tennessee Department of Education, 2016b).

Sample

The population used in this research was Tennessee public school teachers in grades three through twelve who taught a tested subject for the 2014-2015 school year in school systems that used either the TEAM or TIGER evaluation model. The sample for this study consisted of 230 TIGER teachers in three TIGER school systems and 2,839 TEAM teachers in nine TEAM school

systems. During the 2014-2015 school year, 12 school systems used the TIGER model. Three TIGER school systems agreed to participate and were included in the sample for this study. During the 2014-2015 school year, 127 school systems used the TEAM model (Tennessee Department of Education, 2016b), and nine TEAM school systems agreed to participate. For each of the participating school systems, all teachers who taught a tested subject in grades three through twelve were included in the sample.

Data Collection

A request was made to the Institutional Review Board (IRB) for approval to collect TVAAS data and final observation scores from the individual school systems. The IRB concluded that the proposed research did not meet the FDA or DHHS definition of research involving human subjects, so IRB approval was not required. Data for this study were collected from each individual school system. The researcher contacted the director of schools for each school system to request the data and to inform him or her as to the purpose of the study. The researcher followed the protocol designated by each school system in order to collect the data. Data were e-mailed to the researcher by the school system contact in a spreadsheet with teacher and school system names removed to ensure anonymity and confidentiality. The data provided was a numbered list with each teacher's final observation score and TVAAS score together. Participants were given assurance that no identifying school system information or names would be shared in order to protect privacy and to maintain confidentiality.

Data Analysis

In order to address Research Questions 1 and 2, independent samples t-tests were conducted to determine whether observation scores or TVAAS scores differed for teachers

evaluated with the TEAM model and teachers evaluated with the TIGER model. The population for this study was Tennessee public school teachers in grades three through twelve who taught state-tested subjects in 2014-2015 in school systems that used either the TEAM or TIGER teacher evaluation models. The dependent variable was teacher effectiveness as measured by TVAAS growth scores and final observation scores for the 2014-2015 school year. Individual TVAAS growth and observation scores were averaged to find a mean TVAAS and observation score for both TIGER and TEAM school systems. The means were used in an independent samples t-test to investigate differences in teacher effectiveness (Green & Salkind, 2011).

To address Research Questions 3 and 4, Pearson correlation coefficients were used to determine the relationship between observation scores and value-added growth scores for both teachers in the TEAM model and the TIGER model. The Statistical Package for the Social Sciences (SPSS) was used for the data analysis. All data were analyzed at the 0.05 level of significance.

Ch**apter Summary**

Chapter 3 describes the research design, research questions and corresponding null hypotheses, instrumentation, sample, data collection, and data analysis for this research study. SPSS was used to calculate the results of the relationship between teacher evaluation model and teacher effectiveness. The population for this study was Tennessee teachers in grades three through eight who taught tested subjects in school systems that either used the TEAM or TIGER evaluation models. The research addressed four research questions and tested four corresponding null hypotheses. Chapter 4 provides summaries of the data analysis for each research question.

CHAPTER 4

FINDINGS

The purpose of this study was to compare measures of teacher effectiveness for two different Tennessee teacher evaluation models: TEAM and TIGER. The measures of teacher effectiveness that were used in this study were a teacher's final observation score from the TEAM rubric and a teacher's individual value-added growth (TVAAS) score. The relationship between observation scores and TVAAS scores was also compared between the two different evaluation models. Data was gathered from the 2014-2015 school year. The sample for this study consisted of 2,839 TEAM teachers from 9 TEAM school systems and 230 TIGER teachers from 3 TIGER systems. Independent-samples t-tests were conducted to investigate any differences in observation scores and value-added growth scores (TVAAS) between teachers who were evaluated with the TEAM model and teachers who were evaluated with the TIGER model. Person correlations were computed to look at the relationship between observation scores and TVAAS for each of the evaluation models. In this chapter, data analyses were completed to investigate four research questions and four corresponding null hypotheses.

Research Question 1

Is there a significant difference in overall teacher effectiveness as measured by TVAAS between teachers who are evaluated with the TEAM model and teachers evaluated with the TIGER model?

H_{01}: There is no significant difference in overall teacher effectiveness as measured by TVAAS between teachers who are evaluated with the TEAM model and teachers evaluated with the TIGER model.

An independent-samples t-test was conducted to evaluate whether the mean individual value-added growth score (TVAAS) differed between teachers who were evaluated with the TEAM evaluation model and teachers who were evaluated with the TIGER evaluation model. The TVAAS score was the test variable and the grouping variable was the TEAM or TIGER evaluation model. The test was significant, $t(3067) = 4.267$, $p < 0.001$. Therefore, the null hypothesis was rejected. Teachers who were evaluated with the TIGER model ($M = 3.744$, $SD = 1.402$) tended to have higher TVAAS scores than teachers who were evaluated with the TEAM model ($M = 3.320$, $SD = 1.453$). The 95% confidence interval for the difference in means was 0.229 to 0.619. Cohen's d was calculated to be 0.293, which indicated a small to medium effect size. Figure 1 shows the distributions for the two groups. In summary, teachers who were evaluated with the TIGER model tended to have significantly higher TVAAS scores than teachers who were evaluated with the TEAM model.

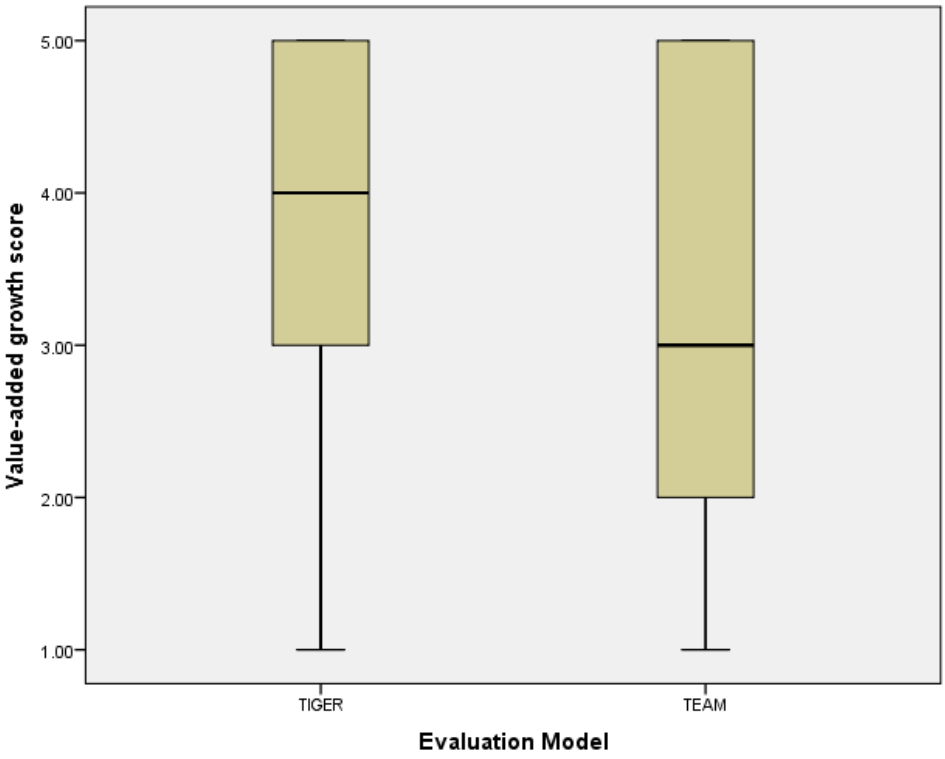

Figure 1: TVAAS Scores for TEAM and TIGER Teachers

Research Question 2

Is there a significant difference in final principal observation scores between teachers who are evaluated with the TEAM model and teachers evaluated with the TIGER model?

H_{02}: There is no significant difference in final principal observation scores between teachers who are evaluated with the TEAM model and teachers evaluated with the TIGER model.

An independent-samples t-test was conducted to evaluate whether the mean final observation score differed for teachers who were evaluated with the TEAM model and teachers evaluated with the TIGER model. The final observation score was the test variable and the grouping variable was the TEAM or TIGER evaluation model. The test was significant, $t(3067) = 7.782$, $p < 0.001$. Therefore, the null hypothesis was rejected. Teachers who were evaluated with the TIGER model ($M = 4.174$, $SD = 0.402$) tended to have a higher final observation score than those teachers who were evaluated with the TEAM model ($M = 3.878$, $SD = 0.565$). The 95% confidence interval for the difference in means was 0.221 to 0.370. Cohen's d was calculated to be 0.535, which indicated a medium effect size. Figure 2 shows the distributions for the two groups. In summary, teachers who were evaluated with the TIGER model tended to have significantly higher final observation scores than teachers who were evaluated with the TEAM model.

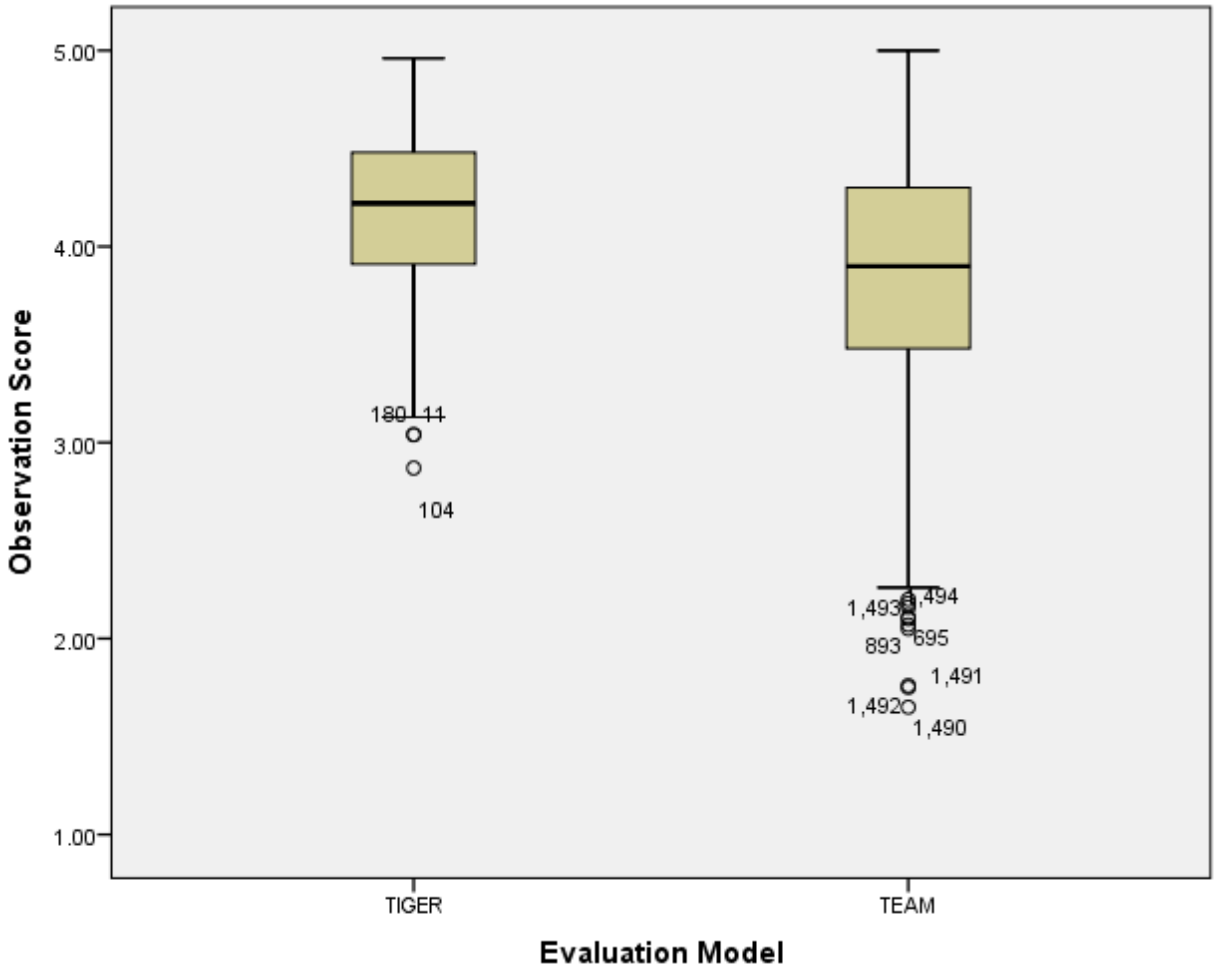

Figure 2: Observation Scores for TEAM and TIGER Teachers.

° Indicates Extreme Value

Research Question 3

Is there a significant relationship between teachers' value-added growth scores (TVAAS) and teachers' final observation scores for teachers who are evaluated with the TEAM model?

H_{03}: There is no significant relationship between teachers' value-added growth scores (TVAAS) and teachers' final observation scores for teachers who are evaluated with the TEAM model.

A Pearson correlation coefficient was computed to test the relationship between TVAAS scores and final observation scores for teachers who were evaluated with the TEAM model. The results of the analysis revealed a significantly positive relationship between final observation scores (M = 3.878, SD = 0.565) and TVAAS scores (M = 3.320, SD = 1.453) and a statistically significant correlation [$r(2837) = 0.359$, $p < 0.001$]. As a result of the analysis, the null hypothesis was rejected. In general, the results suggest that teachers who were evaluated with the TEAM model who tended to have higher final observation scores also tended to have higher TVAAS scores.

Research Question 4

Is there a significant relationship between teachers' value-added growth scores (TVAAS) and teachers' final observation scores for teachers who are evaluated with the TIGER model?

H_{04}: There is no significant relationship between teachers' value-added growth scores (TVAAS) and teachers' final observation scores for teachers who are evaluated with the TIGER model.

A Pearson correlation coefficient was computed to test the relationship between TVAAS scores and final observation scores for teachers who were evaluated with the TIGER model. The results of the analysis revealed a significantly positive relationship between final observation scores (M = 4.174, SD = 0.402) and TVAAS scores (M = 3.744, SD = 1.402) and a statistically significant correlation [$r(228) = 0.531$, $p < 0.001$]. As a result of the analysis, the null hypothesis was rejected. In general, the results suggest that teachers who were evaluated with the TIGER model who tended to have higher final observation scores also tended to have higher TVAAS scores.

Chapter Summary

In this chapter, observation scores and TVAAS scores were compared between the TEAM and TIGER Tennessee teacher evaluation models. The relationship between TVAAS scores and final observation scores were also analyzed and presented. There were four research questions and four corresponding null hypotheses. TVAAS data and observation scores were collected from 12 school systems representing 3,069 teachers. Nine of the school systems used the TEAM model for the 2014-2015 school year, and three of the school systems used the TIGER model for the 2014-2015 school year.

In analyzing the data, the researcher found a significant difference between both observation scores and TVAAS scores between teachers who were evaluated with the TEAM model and teachers who were evaluated with the TIGER model. TIGER teachers tended to have higher final observation scores as well as higher TVAAS scores. The relationship between TVAAS scores and final observation scores for both models was found to be significant and positive. The correlation between TVAAS scores and observation scores tended to be stronger for the TIGER evaluation scores.

CHAPTER 5

SUMMARY OF FINDINGS, RECOMMENDATIONS, AND CONCLUSIONS

This chapter provides a summary of findings, recommendations for practice, recommendations for future research, and conclusions. The purpose of this study was to compare measures of teacher effectiveness for teachers who were evaluated with the Tennessee Educator Acceleration Model (TEAM) and the Teacher Instructional Growth for Effectiveness and Results (TIGER) model. The measures of teacher effectiveness that were used were a teacher's final observation score and a teacher's individual value-added growth score (TVAAS). The correlation between final observation scores and value-added scores were compared between the two different teacher evaluation models. The sample for this study included 230 TIGER teachers and 2,839 TEAM teachers.

Summary of Findings

The analysis of data for this study centered on four research questions and four corresponding null hypotheses. Independent-samples t-tests were conducted for Research Questions 1 and 2 to help determine if there was a significant difference between final observation scores and TVAAS scores between the TEAM and TIGER evaluation models. Pearson correlation coefficients were computed for Research Questions 3 and 4 to help determine the relationship between TVAAS and final observation scores for each of the evaluation models.

Research Question 1

Is there a significant difference in overall teacher effectiveness as measured by TVAAS between teachers who are evaluated with the TEAM model and teachers evaluated with the TIGER model?

The purpose of this research question was to evaluate any difference in TVAAS scores between teachers who were evaluated with the TEAM model and teachers who were evaluated with the TIGER model. The mean TVAAS score for teachers who were evaluated with the TIGER model was significantly higher than the scores of teachers who were evaluated with the TEAM model. Results of this analysis were similar to the data presented in 2012 from the Tennessee Department of Education, in which it was indicated that a higher percentage of TIGER teachers earned a TVAAS score of a five, the highest score possible, when compared to TEAM teachers (Tennessee Department of Education, 2012).

Research Question 2

Is there a significant difference in final principal observation scores between teachers who are evaluated with the TEAM model and teachers evaluated with the TIGER model?

The purpose of this research question was to evaluate any difference in final observation scores between teachers who were evaluated with the TIGER model and teachers who were evaluated with the TEAM model. Teachers who were evaluated with the TIGER model tended to have higher observation scores than teachers who were evaluated with the TEAM model. Results of this analysis also echoed information shared in a 2012 report from the Tennessee Department of Education, which indicated that a higher percentage of TIGER teachers earned either a four or

five as a final observation score when compared with TEAM teachers (Tennessee Department of Education, 2012).

Research Question 3

Is there a significant relationship between teachers' value-added growth scores (TVAAS) and teachers' final observation scores for teachers who are evaluated with the TEAM model?

The purpose of this research question was to evaluate the relationship between individual value-added growth (TVAAS) scores and final observation scores for teachers who were evaluated with the TEAM model. Results indicated a significantly positive relationship between observation scores and TVAAS. TEAM teachers who tended to have higher final observation scores also tended to have higher TVAAS scores. The results of this research support the findings of Gallagher (2004), Garret and Steinberg (2015), Gill et al. (2016), Kane et al. (2010), and Mason (2015).

Research Question 4

Is there a significant relationship between teachers' value-added growth scores (TVAAS) and teachers' final observation scores for teachers who are evaluated with the TIGER model?

The purpose of this research question was to evaluate the relationship between individual value-added growth (TVAAS) scores and final observation scores for teachers who were evaluated with the TIGER model. Results indicated a significant positive relationship between observation scores and TVAAS. TIGER teachers who tended to have higher final observation scores also tended to have higher TVAAS scores. There was a 0.172 difference in correlation

coefficients between TEAM and TIGER teachers indicating the correlation between observation scores and value-added scores was stronger for TIGER teachers.

Recommendations for Practice

Based on the results of this study, the following recommendations for practice are suggested:

1. The results of this study indicated that there was a significant difference in measures of teacher effectiveness between TEAM and TIGER teachers. Both the mean observation score and TVAAS score were higher for TIGER teachers. One of the differences between the two models is how final observation scores are computed. With the TEAM model, the final score is an average of all of the observations throughout the year. With the TIGER model, teachers are allowed to provide evidence in the form of artifacts to support his or her expertise in an area of the rubric. Walkthrough observations from administrators inform TIGER teachers as to areas where there is improvement or the need to provide evidence. This is based on the idea of formative assessment. Based on this concept, it is recommended that all administrators who perform observations provide more and consistent feedback to teachers to help them improve their practice.
2. The TEAM rubric, which both the TEAM and TIGER models use, is based on research as to what effective teachers do in their classrooms. Teachers should be offered in-depth training in how to understand the TEAM rubric as well as training as to what effective teaching looks like in each of the domains of the rubric.

3. Although this research has shown a significantly positive relationship between observation scores and value-added scores, this relationship could be strengthened. Administrators should continuously be working to improve in the area of performing teacher observations. Continued training in the use of the rubric as well as ways to identify effective practices in the classroom should be the focus of professional development for administrators.
4. Teachers should also be provided professional development regarding how to use value-added scores to improve practice.
5. School systems should evaluate how well the chosen evaluation model has been implemented. Administrators should engage in an in-depth look at observations and how they are performed in the school system.
6. School systems should spend time investigating all of the evaluation model options to see if there are any components that make one option a better fit for that school system.
7. Administrators should provide professional development to teachers in areas that research has shown to impact teacher effectiveness such as classroom management (Decristan et al., 2015; Grant et al., 2013, Kane et al., 2011, Stronge et al, 2011; Tyler et al., 2015), building positive relationships (Hattie, 2012; Mangiante, 2011; Stronge et al., 2011), giving constructive feedback (Grossman et al., 2010; Mangiante, 2011), and setting high expectations (Hattie, 2012; Mangiante, 2011; Stronge et al., 2008)

Recommendations for Future Research

The results of this study indicate that there are statistically significant differences in measures of teacher effectiveness between TEAM and TIGER teachers. Results also indicate that

the correlation between final observation scores and TVAAS scores is different between the two different evaluation models. The following are recommendations for future research, which may further support the ideas in this study as well as add to the body of research regarding effective teacher evaluation models.

1. This study should be replicated with a larger sample of teacher evaluation data from TEAM and TIGER school systems. It is important to see if similar results are found with a larger sample.
2. This study should be replicated longitudinally to see if the results of this study hold up over a longer period of time.
3. This study was limited to teachers who taught a state-tested subject due to the fact that these teachers had an individual value-added growth score. There are many teachers who were not included in this study because they taught subjects that were not tested. The state has begun implementing portfolios as a way to evaluate these teachers. Currently, there are portfolio options for teachers who teach fine arts, first grade, physical education, pre-kindergarten, kindergarten, and foreign languages (Tennessee Department of Education, 2016b). Therefore, future study is recommended to compare portfolio data to see if there is a significant difference in teacher effectiveness as measured through teacher portfolios for TEAM and TIGER teachers.
4. With the 2016-2107 school year, the Tennessee Department of Education began using a new standardized test to measure student learning. Therefore, this study should be replicated using teacher evaluation data based on these new assessments.

5. Based on the research, there are differences in the TEAM and TIGER models in the structure of the teacher observations, the use of artifacts as evidence for teacher effectiveness, and final observation scoring. Future research should be conducted to investigate how well the different components of these different evaluation models are actually implemented. Are school systems using the evaluation models as they are intended and designed?
6. Research has indicated that there are many factors that could impact teacher effectiveness such as years of experience (Chingos & Peterson, 2011; Munoz et al., 2011) and highest level of education attained (Darling-Hammond, 2000). This study could be replicated but control for level of education and years of experience.
7. This study was delimited to the TEAM and TIGER Tennessee teacher evaluation models because they are the most used in the state of Tennessee, but there are three other Tennessee approved alternate evaluation models. Therefore, future research should compare measures of teacher effectiveness between all of the evaluation model options.
8. An in-depth qualitative research study should be performed to further investigate differences between the two evaluation models incorporating teacher interviews as well as evaluation artifacts.

Conclusions

Teacher evaluations have gone through many changes and phases over time. Currently, there is a significant mandate for accountability. Part of many teacher evaluation models include some value-added component, which represents how much of an impact a teacher has had on student growth and learning. With this mandate for accountability, it is important to provide

teachers with as many resources to help them be successful in the classroom. The goal of this research study was to compare teacher effectiveness measures between the TEAM and TIGER evaluation models. In having a greater understanding of evaluation models and their components, administrators can help teachers to understand their performance and ways to improve. In this study, both observation scores and individual value-added growth scores were compared between the TEAM and TIGER evaluation models. For both effectiveness measures, there was a statistically significant difference between the two models. TIGER teachers had a higher mean observation score as well as a higher mean TVAAS score. This could indicate that perhaps some of the differences that distinguish the two models such as the focus of observations, the use of evidence and walkthroughs, as well as how the final score is calculated could provide information to teachers to help them improve and grow as teachers. With further research into why there are differences, administrators could develop methods and practices to help support teachers.

This study also examined the relationship between value-added scores and observation scores between the two models. Both the TEAM and TIGER model had a positive, significant, correlation between TVAAS and observation scores. The TIGER model scores had a stronger correlation, again suggesting that some component of the TIGER model that differentiates it from the TEAM model may be part of the reason. Consequences of this study indicate that it is worthwhile to study this topic more in-depth, to look for possible reasons why the TIGER teachers had higher mean scores on measures of teacher effectiveness as well as a stronger correlation between value-added and observation scores.

REFERENCES

Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics*, 25(1), 95-135. Retrieved October 29, 2015 from http://www.jstor.org/stable/10.1086/508733

Amrein-Beardsley, A. (2008). Methodological concerns about the education value-added assessment system. *Educational Researcher*, 37(2), 65-75. Retrieved June 6, 2016 from https://login.iris.etsu.edu:3443/login?url=http://search.proquest.com.iris.etsu.edu:2048/docview/216901808?accountid=10771

Baker, A. P., & Xu, D. (1995). *The measure of education: A review of the Tennessee value added assessment system*. Nashville, TN: Office of Educational Accountability. Retrieved June 20, 2016 from http://files.eric.ed.gov/fulltext/ED388697.pdf

Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics*, 29(1), 37-65. Retrieved June 6, 2016 from http://journals.sagepub.com.iris.etsu.edu:2048/doi/pdf/10.3102/10769986029001037

Bell, C. A., Qi, Y., Croft, A. J., Leusner, D., McCaffrey, D. F., Gitomer, D. H., & Pianta, R. C. (2014). Improving observational score quality: Challenges in observer thinking. In T. Kane, K. Kerr, & R. Pianta (Eds.), *Designing teacher evaluation systems* (pp. 50-97). San Francisco, CA: Josey-Bass.

Bobbit, F. (1918). *The curriculum*. Boston, MA: Houghton Mifflin.

Bogart, C.,D. (2013). *Teacher evaluation and classroom practice: Teacher perceptions in northeast Tennessee* (Doctoral dissertation). Retrieved September 8, 2015 from http://search.proquest.com/docview/1442476310/

Bryant, C. B. (2013). *Perceptions of Tennessee school principals about the Tennessee Educator Acceleration Model (TEAM)* (Doctoral dissertation). Retrieved September 8, 2015 from http://search.proquest.com.iris.etsu.edu:2048/docview/1371990887/

Callahan, R. E. (1962). *Education and the cult of efficiency*. Chicago, IL: The University of Chicago Press.

Chingos, M. M., & Peterson, P. (2011). It's easier to pick a good teacher than to train one: Familiar and new results on the correlates of teacher effectiveness. *Economics of Education Review*, 30(3), 449-465. Retrieved May 25, 2016 from http://www.sciencedirect.com.iris.etsu.edu:2048/science/article/pii/S0272775710001755

Cogan, M. L. (1973). *Clinical supervision*. Boston, MA: Houghton Mifflin.

Creswell, J. W. (2002). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Upper Saddle River, NJ: Pearson.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, CA: SAGE.

Cubberly, E. P. (1916). *Public school administration: A statement of the fundamental principles underlying the organization and administration of public education*. Boston, MA: Houghton Mifflin.

Danielson, C. (2007). *Enhancing professional practice: A framework for teaching*. Alexandria, VA: ASCD.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives*, 8(1), 1-44. Retrieved November 11, 2015 from http://dx.doi.org/10.14507/epaa.v8n1.2000

Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *Phi Delta Kappan*, 93(6), 8-15. Retrieved October 13, 2015 from http://go.galegroup.com.iris.etsu.edu:2048/ps/i.do?&id=GALE|A284016616&v=2.1&u=tel_a_etsul&it=r&p=AONE&sw=w&authCount=1#

Davis, J.B. (2014). *The relationship between the growth score and the overall TEAM observation rating for teachers in Tennessee* (Doctoral dissertation). Retrieved on September 13, 2015 from http://dc.etsu.edu/etd/2432

Decristan, J., Klieme, E., Kunter, M., Hochweber, J., Buttner, G., Fauth, B.,…Hardy, I. (2015). Embedded formative assessment and classroom process quality: How do they interact in promoting science understanding? *American Educational Research Journal*, 52(6), 1133-1159. Retrieved July 17, 2016 from 10.3102/0002831215596412

Ehlert, M. W., Pepper, M. J., Parsons, E. S., Burns, S. F., & Springer, M., G. (2013). *Educator evaluation in Tennessee: Initial findings from the 2013 first to the top survey*. Retrieved October 1, 2015 from https://peabody.vanderbilt.edu/research/tnedresearchalliance/files/Educator_Evaluation_in_Tennessee_Initial_Findings_from_the_2013_First_to_the_Top_Survey13.pdf

Every Student Succeeds Act of 2015, Pub. L. No. 114-95 § 114, Stat. 1177 (2015).

Ferguson, R. F., & Danielson, C. (2014). How framework for teaching and tripod 7Cs evidence distinguish key components of effective teaching. In T. J. Kane, K. A. Kerr & R. C. Pianta (Eds.), *Designing teacher evaluation systems* (pp. 98-143). San Francisco, CA: Jossey-Bass.

Figlio, D., & Loeb, S. (2011). School accountability. In E.A. Hanushek, S. Machin & L. Woessman (Eds.), *Handbooks of Economics: Economics of Education* (Vol.3, pp. 383-421). Amsterdam, Netherlands: North-Holland.

Furtwengler, C. (1985). Tennessee's career ladder plan: They said it couldn't be done! *Educational Leadership*, 43(3), 50-56. Retrieved June 7, 2016 from http://www.ascd.org/ASCD/pdf/journals/ed_lead/el_198511_furtwengler.pdf

Gallagher, H. A., (2004). Vaughn Elementary's innovative teacher evaluation system: Are teacher evaluation scores related to growth in student achievement? *Peabody Journal of Education*, 79(1), 79-107. Retrieved October 19, 2015 from https://login.iris.etsu.edu:3443/login?url=https://search-proquest-com.iris.etsu.edu:3443/docview/62124576?accountid=10771

Garret, R., & Steinberg, M. P. (2015). Examining teacher effectiveness using classroom observation scores: Evidence from the randomization of teachers to students. *Educational Evaluation and Policy Analysis*, 37(2), 224-242. Retrieved October 15, 2016 from http://journals.sagepub.com.iris.etsu.edu:2048/doi/abs/10.3102/0162373714537551

Gill, B., Shoji, M., Coen, T., & Place, K. (2016). *The content, predicative power, and potential bias in five widely used teacher observation instruments* (REL 2017-191). Retrieved March 7, 2017 from Institute of Education Sciences U.S. Department of Education website: https://ies.ed.gov/ncee/edlabs/regions/midatlantic/pdf/REL_2017191.pdf.

Goldhammer, R. (1969). *Clinical supervision: Special methods for the supervision of teachers*. New York, NY: Holt, Rinehart, and Winston.

Goldring, E., Grissom, J. A., Rubin, M., Neumerski, C. M., Cannata, M., Drake, T., & Schuermann, P. (2015). Make room value added: Principals' human capital decisions and the emergence of teacher observation data. *Educational Researcher*, 44(2), 96-104**.** Retrieved October 18, 2015 from 10.3102/0013189X15575031

Grant, L. W., Stronge, J. H., & Xu, X. (2013). A cross-cultural comparative study of teacher effectiveness: Analyses of award winning teachers in the United States and China. *Educational Assessment, Evaluation and Accountability*, 25(3), 251-276. Retrieved March 25, 2016 from http://dx.doi.org.iris.etsu.edu:2048/10.1007/s11092-013-9170-1

Green, S. B., & Salkind, N. J. (2011). *Using SPSS for Windows and Macintosh: Analyzing and understanding data* (6th ed.). Boston, MA: Pearson.

Grossman, P., Cohen, J., Ronfeldt, M., & Brown, L. (2014). The test matters: The relationship between classroom observation scores and teacher value added on multiple types of assessment. *Educational Researcher*, 43(6), 293-303. Retrieved October 13, 2015 from http://journals.sagepub.com.iris.etsu.edu:2048/doi/abs/10.3102/0013189X14544542

Grossman, P., Loeb, S., Cohen, J., Hammerness, K., Wyckoff, K., Boyd, D., & Lankford, H. (2010). *Measure for measure: The relationship between measures of instructional practice in middle school English language arts and teachers' value-added scores* (NBER Working Paper No. 16015). Retrieved November 7, 2015 from National Bureau of Economic Research website: http://www.nber.org/papers/w16015

Hallinger, P., Heck, R. H., & Murphy, J. (2014). Teacher evaluation and school improvement: An analysis of the evidence. *Educational Assessment, Evaluation and Accountability*, 26(1), 5-28. Retrieved September 9, 2015 from http://dx.doi.org/10.1007/s11092-013-9179-5

Hanushek, E. A. (1992). The trade-off between child quantity and quality. *Journal of Political Economy*, 100(1), 84-117. Retrieved July 8, 2016 from http://www.jstor.org.iris.etsu.edu:2048/stable/2138807

Harris, D. N., Ingle, W. K., & Rutledge, S. A. (2014). How teacher evaluation methods matter for accountability: A Comparative analysis of teacher effectiveness ratings by principals and teacher value-added measures. *American Educational Research Journal*, 51(1), 73-112. Retrieved September 8, 2015 from 10.3102/0002831213517130

Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. New York, NY: Routledge.

Heneman, H. G., Milanowski, A., Kimball, S. M., & Odden, A. (2006). Standards-based teacher evaluation as a foundation for knowledge- and skill-based pay. White paper retrieved November 1, 2015 from http://www.cpre.org/sites/default/files/policybrief/885_rb45.pdf

Hill, H. C., & Grossman, P. (2013). Learning from teacher observations: Challenges and opportunities posed by new teacher evaluation systems. *Harvard Educational Review*, 83(2), 371-384. Retrieved September 9, 2015 from https://search-proquest-com.iris.etsu.edu:3443/docview/1399327195/fulltext/D544955DA6F425DPQ/1?accountid=10771

Hoole, C. (1659). *Scholastick discipline or the way of ordering a grammar schoole*. London, England: J.T.

Hull, J. (2013). *Trends in teacher evaluation: How states are measuring teacher performance*. Retrieved July 15, 2016 from the Center for Public Education website: http://www.centerforpubliceducation.org/Main-Menu/Evaluating-performance/Trends-in-Teacher-Evaluation-At-A-Glance/Trends-in-Teacher-Evaluation-Full-Report-PDF.pdf

Jacob, B. A.., & Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics*, 26(1), 101-136. Retrieved October 25, 2015 from http://www.jstor.org.iris.etsu.edu:2048/stable/10.1086/522974

Jiang, J. Y., Sporte, S. E., & Luppescu, S. (2015). Teacher perspectives on evaluation reform: Chicago's REACH students. *Educational Researcher*, 44(2), 105-116. Retrieved September 9, 2105 from 10.3102/0013189X15575517

Joe, J. N., McClellan, C. A., & Holtzman, S. L. (2014). Scoring design decisions: Reliability and the length and focus of classroom observations. In T. J. Kane, K. A. Kerr & R. C. Pianta (Eds.), *Designing teacher evaluation systems* (pp. 415-443). San Francisco, CA: Jossey-Bass.

Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2011). Identifying effective classroom practices using student achievement data. *The Journal of Human Resources*, 46(3), 588-613. Retrieved October 16, 2015 from http://search.ebscohost.com.iris.etsu.edu:2048/login.aspx?direct=true&db=ssf&AN=508447162&site=ehost-live

Kelly, D., Xie, H., Nord, C. W., Jenkins, F., Chan, J. Y., & Kastberg, D. (2013). *Performance of U.S. 15-year-old students in mathematics, science, and reading literacy in an international context: First look at PISA 2012* (NCES Report 2014-024). Retrieved July 20, 2016 from National Center for Education Statistics website: http://nces.ed.gov/pubs2014/2014024rev.pdf

Kimball, S. M., & Milanowski, A. (2009). Examining teacher validity and leadership decision making within a standards-based evaluation system. *Educational Administration Quarterly*, 45(1), 34-70. Retrieved October 12, 2015 from 10.1177/0013161X08327549

Kimball, S. M., White, B., Milanowski, A. T., & Borman, G. (2004). Examining the relationship between teacher evaluation and student assessment results in Washoe county. *Peabody Journal of Education*, 79(4), 54078. Retrieved October 15, 2015 from https://login.iris.etsu.edu:3443/login?url=https://search-proquest-com.iris.etsu.edu:3443/docview/62124603?accountid=10771

Kupermintz, H. (2003). Teacher effects and teacher effectiveness: A validity investigation of the Tennessee value added assessment system. *Educational Evaluation and Policy Analysis*, 25(3), 287-298. Retrieved October 20, 2015 from https://search-proquest-com.iris.etsu.edu:3443/docview/197237045?OpenUrlRefId=info:xri/sid:primo&accountid=10771

Mangiante, E. M. (2011). Teachers matter: Measures of teacher effectiveness in low-income minority schools. *Educational Assessment, Evaluation and Accountability*, 23(1), 41-63. Retrieved June 15, 2016 from http://dx.doi.org.iris.etsu.edu:2048/10.1007/s11092-010-9107-x

Martin, M. O., Mullis, I.V., Foy, P., & Stanco, G. M. (2012). *TIMSS 2011 international results in science*. Retrieved July 20, 2016 from the TIMSS and PIRLS International Study Center website: http://www.bc.edu/content/dam/files/research_sites/timssandpirls/timss2011/downloads/T11_IR_Science_FullBook.pdf

Marzano, R. J., Frontier, T., & Livingston, B. (2011). *Effective supervision: Supporting the art and science of teaching*. Alexandria, VA: ASCD.

Mason, J.M. (2015). *The relationship between teacher observation scores in English II and III for thinking and problem solving and student growth and achievement* (Doctoral dissertation). Retrieved November 7, 2015 from http://dc.etsu.edu/etd/2432

McMillan, J. H., & Schumacher, S. (2010). *Research in education: Evidence-based inquiry* (7th ed.). Boston, MA: Pearson.

Mullis, I. V., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 international results in mathematics*. Retrieved July 20, 2016 from the TIMSS and PIRLS International Study Center website: https://timssandpirls.bc.edu/timss2011/downloads/T11_IR_Mathematics_FullBook.pdf

Munoz, M. A., & Chang, F. C. (2007). The elusive relationship between teacher characteristics and student academic growth: A longitudinal multilevel model for change. *Journal of*

Personnel Evaluation in Education, 20(3), 147-164. Retrieved June 16, 2016 from http://dx.doi.org/10.1007/s11092-008-9054-y

Munoz, M. A., Prather, J. R., & Stronge, J. H. (2011). Exploring teacher effectiveness using hierarchical linear models: Student- and classroom-level predictors and cross-year stability in elementary school reading. *Planning and Changing*, 42(3-4), 241-273. Retrieved June 1, 2016 from https://search-proquest-com.iris.etsu.edu:3443/docview/1508228040?OpenUrlRefId=info:xri/sid:primo&accountid=10771

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform: A report to the nation and the Secretary of Education, United States Department of Education*. Washington, DC: The Commission.

New Teacher Center. (2013). *2013 TELL Tennessee survey: Teaching conditions across evaluation models*. Retrieved November 1, 2015 from https://telltennessee.org/uploads/File/TN13_alt%20eval%20model%20brief_2013.8.23_final%20to%20client.pdf

No Child Left Behind Act of 2001, Pub. L. No. 107-110 § 115, Stat, 1425 (2002).

Pajak, E. (2003). *Honoring diverse teaching styles: A guide for supervisors*. Alexandria, VA: ASCD.

Papay, J. P. (2012). Refocusing the debate: Assessing the purposes and tools of teacher evaluation. *Harvard Educational Review*, 82(1), 123-141. Retrieved September 9, 2015 from https://search-proquest-com.iris.etsu.edu:3443/docview/963359551?OpenUrlRefId=info:xri/sid:primo&accountid=10771

Park, Y. S., Chen, J., & Holtzman, S. L. (2014). Evaluating efforts to minimize rater bias in scoring classroom observations. In T. Kane, K. Kerr, & R. Pianta (Eds.), *Designing teacher evaluation systems* (pp. 383-414). San Francisco, CA: Jossey-Bass.

Pepper, M. J., Ehlert, M. W., Parsons, E. S., Stahlheber, S. W., & Burns, S. F. (2015). *Educator evaluation in Tennessee: Findings from the 2014 first to the top survey.* Retrieved October 22, 2015 from Tennessee Consortium on Research, Evaluation & Development website https://peabody.vanderbilt.edu/research/tnedresearchalliance/files/2014_FirstToTheTopSurveyReport.pdf

Polikoff, M. S. (2014). Does the test matter? Evaluating teachers when tests differ in their sensitivity to instruction. In T. Kane, K. Kerr, & R. Pianta (Eds.), *Designing teacher evaluation systems* (pp. 278-302). San Francisco, CA: Jossey-Bass.

Sanders, W. L., & Horn, S. P. (1994). The Tennessee value-added assessment system (TVAAS): mixed-model methodology in educational assessment. *Journal of Personnel Evaluation in Education*, 8(3), 299-311. Retrieved July 18, 2106 from http://s3.amazonaws.com/academia.edu.documents/27368068/the_tennessee_value-added_assessment_system-_mixed-model_methodology_in_ed_assessment.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1499804864&Signature=FPUyNhSsgrIEGji2jOjOA3jxAY8%3D&response-content-disposition=inline%3B%20filename%3DThe_Tennessee_value-added_assessment_sys.pdf

Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement* (Research Report No. R11-0435-02-001-97). Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.

Sanders, W. L., Saxton, A. M., Schneider, J. F., Dearden, B. L., Wright, S. P., & Horn, S. P. (1994). *Effects of building change on indicators of student academic growth*. June 6, 2016 from http://www.cgp.upenn.edu/pdf/Sanders_et_al_-_Effects_of_Building_Change.pdf

SAS Institute. (2015). *Tennessee Department of Education: Technical documentation for 2015 TVAAS analyses*. Retrieved October 10, 2105 from https://www.tn.gov/assets/entities/education/attachments/tvaas_technical_documentation_2015.pdf

Schenker, J. D., & Rumrill, P. D. (2004). Causal-comparative research designs. *Journal of Vocational Rehabilitation*, 21(3), 117-121. Retrieved July 13, 2016 from http://search.ebscohost.com.iris.etsu.edu:2048/login.aspx?direct=true&db=ccm&AN=106626286&site=ehost-live

SCORE. (2014). *Measuring student growth in Tennessee: Understanding TVAAS*. Retrieved June 10, 2016 from http://tnscore.org/wp-content/uploads/2014/10/TVAAS-Memo.png

Shinkfield, A. J., & Stufflebeam, A. J. (1995). *Teacher evaluation: Guide to effective practice*. Norwell, MA: Kluwer.

Shuls, J. V., & Trivitt, J. R. (2015). Teacher effectiveness: An analysis of licensure screens. *Educational Policy*, 29(4), 645-675. Retrieved June 13, 2016 from http://journals.sagepub.com.iris.etsu.edu:2048/doi/abs/10.1177/0895904813510777

Smith, M. J. (2004). *The Education Improvement Act: A progress report*. Retrieved July 11, 2016 from Tennessee Comptroller of the Treasury website: http://www.comptroller.tn.gov/repository/RE/educimproveact.pdf

Staiger, D. O., & Kane, T. J. (2014). Making decisions with imprecise performance measures: The relationship between annual student achievement gains and a teacher's career value added. In T. Kane, K. Kerr, & R. Pianta (Eds.), *Designing teacher evaluation systems* (pp.144-169). San Francisco, CA: Josey-Bass.

Strong, M., Gargani, J., & Hacifazlioglu. (2011). Do we know a successful teacher when we see one? Experiments in the identification of effective teachers. *Journal of Teacher Education*, 62(4), 367-382. Retrieved May 24, 2016 from 10.1177/0022487110390221

Stronge, J. H., Ward, T. J., & Grant, L. W. (2011). What makes good teachers good? A cross-case analysis of the connection between teacher effectiveness and student achievement. *Journal of Teacher Education*, 62(4), 339-355. Retrieved June 10, 2106 from http://journals.sagepub.com.iris.etsu.edu:2048/doi/abs/10.1177/0022487111404241

Stronge, J. H., Ward, T. J., Tucker, P. D., & Hindman, J. L. (2008). What is the relationship between teacher quality and student achievement? An exploratory study. *Journal of*

Personnel Evaluation in Education, 20(3), 165-184. Retrieved June 10, 2106 from http://dx.doi.org/10.1007/s11092-008-9053-z

Teacher Instructional Growth for Effectiveness and Results. (2016). *Tiger: A formative model for teacher evaluation*. Retrieved October 12, 2015 from http://tigermodel.net/about/

Tennessee Code Annotated §49-1-603 (2016).

Tennessee Department of Education. (2012). *Teacher evaluation in Tennessee: A report on year 1 implementation*. Retrieved November 15, 2015 from https://www2.ed.gov/programs/racetothetop/communities/tle2-year-1-evaluation-report.pdf

Tennessee Department of Education. (2015). *Teacher and administrator evaluation in Tennessee: A report on year 3 implementation*. Retrieved November 15, 2015 from http://team-tn.org/wp-content/uploads/2013/08/rpt_teacher_evaluation_year_31.pdf

Tennessee Department of Education. (2016a). *Frequently asked questions: 2016-2017 state assessments*. Retrieved July 13, 2016 from http://tn.gov/assets/entities/education/attachments/2016-17_Assessment_FAQ_for_Educators_-_FINAL.PDF

Tennessee Department of Education. (2016b). *TEAM: Tennessee educator acceleration model*. Retrieved October 22, 2015 from http://team-tn.org/

Tennessee First to the Top Act of 2010. (2010). Retrieved June 27, 2016 from http://www.tennessee.gov/sos/acts/106/pub/pc0002EOS.pdf

Tennessee State Board of Education. (1992). *Master plan for Tennessee Schools: Preparing for the twenty-first century*. Retrieved June 27, 2016 from https://eric.ed.gov/?id=ED407703

Tennessee State Board of Education. (2015). *Teacher and principal evaluation policy*. Retrieved November 11, 2105 from https://tn.gov/assets/entities/sbe/attachments/1-30-15-I_C_Teacher_and_Principal_Evlauation_Policy_Attachment.pdf

Tyler, J. H., Taylor, E. S., Kane, T. J., & Wooten, A. L. (2010). Using student performance data to identify effective classroom practices. *American Economic Review*, 100(2), 256-260. Retrieved March 5, 2017 from http://dx.doi.org.iris.etsu.edu:2048/10.1257/aer.100.2.256

United States Department of Education. (2009). *Race to the top program executive summary*. Retrieved November 24, 2015 from https://www2.ed.gov/programs/racetothetop/executive-summary.pdf

United States Department of Education. (2016). *Race to the top fund*. Retrieved November 24, 2015 from http://www2.ed.gov/programs/racetothetop/index.html

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness. *The Education Digest*, 75(2), 31-35. Retrieved October 19, 2015 from https://search-proquest-com.iris.etsu.edu:3443/docview/218196265?OpenUrlRefId=info:xri/sid:primo&accountid=10771

Wetzel, W. A. (1929). Scientific supervision and curriculum building. *The School Review*, 37(2), 117-123. Retrieved July 11, 2016 from http://www.jstor.org/stable/1078902

Wise, A.E., Darling-Hammond, L., McLaughlin, M. W., & Bernstein, H. T. (1984). *Teacher evaluation: A study of effective practices*. (NIE Report R-3139-NIE). Retrieved July 13, 2016 from RAND Corporation website: https://www.rand.org/content/dam/rand/pubs/reports/2006/R3139.pdf

Witte, R. S., & Witte, J. S. (2010). *Statistics* (9th ed.). Hoboken, NJ: John Wiley & Sons.

Wright, R. (2012). *Recent policy changes in Tennessee: Teacher evaluations*. Retrieved June 28, 2016 from Tennessee Comptroller of the Treasury website: http://www.comptroller.tn.gov/Repository/RE/Teacher%20Evaluations.pdf

Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, 11(1), 57-67. Retrieved July 16, 2016 from http://www.bulldogcia.com/Documents/Articles/Wright_Horn_Sanders_1997.pdf

VITA

NICHOLE MORRIS

Education	Central Catholic High School, Toledo, OH 1999
	B.S. Middle Childhood Education, John Carroll University, University Heights, OH, 2003
	M. Ed. Curriculum and Instruction, Bowling Green State University, Bowling Green, OH, 2006
	M.Ed. Educational Media and Educational Technology, East Tennessee State University, Johnson City, TN 2011
	Ed.D. Educational Leadership, East Tennessee State University, Johnson City, TN 2017
Professional Experiences:	7th and 8th Grade Math Teacher, St. Catherine of Sienna Elementary School, Toledo, OH, 2004-2006
	7th and 8th Grade Math Teacher, Heritage Middle School, Middlebury, IN, 2006-2008
	7th and 8th Grade Math Teacher, University School, Johnson City, TN, 2008-2011
	Library Media Specialist, Greeneville Middle School Greeneville, TN, 2011-present